NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

EVALUATION OF THE STATISTICS OF TARGET SPECTRA IN HYPERSPECTRAL IMAGERY (HSI)

by

Joel C. Robertson

September 2000

Thesis Advisor: J. Scott Tyo
Second Reader: Richard C. Olsen

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE September 2000	3. REPORT TYPE AND DATES COVERED Master's Thesis

4. TITLE AND SUBTITLE: Evaluation of the Statistics of Target Spectra in Hyperspectral Imagery(HSI)	5. FUNDING NUMBERS
6. AUTHOR(S) Robertson, Joel C.	

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000	8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.	12b. DISTRIBUTION CODE

13. ABSTRACT *(maximum 200 words)*

The majority of spectral imagery classifiers make a decision based on information from a particular spectrum, often the mean, which best represents the spectral signature of a particular target. It is known, however, that the spectral signature of a target can vary significantly due to differences in illumination conditions, target shape, and target material composition. Furthermore, many targets of interest are inherently mixed, as is the case with camouflaged military vehicles, leading to even greater variability.

In this thesis, a detailed statistical analysis is performed on HYDICE imagery of Davis Monthan Air Force Base. Several hundred pixels are identified as belonging to one of eight target classes and the distribution of spectral radiance within each group is studied. It has been found that simple normal statistics do not adequately model either the total radiance or the single band spectral radiance distributions, both of which can have highly skewed histograms even when the spectral radiance is high. Goodness of fit tests are performed for maximum likelihood normal, lognormal, gamma, and Weibull distributions. It was discovered that lognormal statistics can model the total radiance and many single-band distributions reasonable well, possibly indicative of multiplicative noise features in remotely sensed spectral imagery.

14. SUBJECT TERMS Spectral Imagery, Hyperspectral Imagery, Scene statistics in spectral imagery	15. NUMBER OF PAGES 124
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL

THIS PAGE INTENTIONALLY LEFT BLANK

EVALUATION OF THE STATISTICS OF TARGET SPECTRA IN HYPERSPECTRAL IMAGERY (HSI)

Joel C. Robertson
Lieutenant, United States Navy
B.S., United States Naval Academy, 1994

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN SYSTEMS ENGINEERING

from the

NAVAL POSTGRADUATE SCHOOL
September 2000

ABSTRACT

The majority of spectral imagery classifiers make a decision based on information from a particular spectrum, often the mean, which best represents the spectral signature of a particular target. It is known, however, that the spectral signature of a target can vary significantly due to differences in illumination conditions, target shape, and target material composition. Furthermore, many targets of interest are inherently mixed, as is the case with camouflaged military vehicles, leading to even greater variability.

In this thesis, a detailed statistical analysis is performed on HYDICE imagery of Davis Monthan Air Force Base. Several hundred pixels are identified as belonging to one of eight target classes and the distribution of spectral radiance within each group is studied. It has been found that simple normal statistics do not adequately model either the total radiance or the single band spectral radiance distributions, both of which can have highly skewed histograms even when the spectral radiance is high. Goodness of fit tests are performed for maximum likelihood normal, lognormal, gamma, and Weibull distributions. It was discovered that lognormal statistics can model the total radiance and many single-band distributions reasonably well, possibly indicative of multiplicative noise features in remotely sensed spectral imagery.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

THIS PAGE INTENTIONALLY LEFT BLANK

I.   INTRODUCTION

Hyperspectral imagery (HSI) has been developed as a potentially powerful tool for classifying scenes and identifying targets in remote sensing. With instruments such as the Hyperspectral Digital Imaging Collection Experiment (HYDICE) and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) that measure spectral radiance in over 200 bands in the visible and near infrared portions of the spectrum at each pixel in a scene, the capability exists to identify targets that would not be detectable in lower-spectral resolution images. Many powerful classification strategies have been developed that use information about potential targets and backgrounds. However, the majority of such strategies, including spectral angle mapping (SAM), spectral feature fitting, matched filtering, linear unmixing, and subspace projection techniques are all quasi-deterministic. That is, the techniques are derived based upon certain representative (often mean) spectral signatures. When variations are present, simple assumptions are made about the distribution, and a thresholding operation is performed. While many quasi-deterministic strategies have been shown to be relatively successful, little effort has been dedicated to developing fully stochastic models. Stein (1999) has demonstrated in RADAR detection applications that inclusion of realistic probability models can significantly improve the probability of detection for a given false alarm rate, so it is reasonable to believe that such models could have applications in spectral imagery.

To date, the majority of authors have assumed that HSI distributions could be made up by considering only signals and noise. A best spectral signature was determined based on available data, then all deviations from this were considered to be due to noise,

which is generally assumed to be multi-variate normal in nature. It is known that such a strategy is only an approximation, as two pixels that both contain the same target or material will almost invariably have different spectra, even under the same illumination and atmospheric conditions and in the absence of noise. Basically, this means that the targets in and of themselves are best described by *statistical distributions*, not just the mean spectrum. Approximating the target distributions by the mean spectrum is a first order approximation that matches only the first moment of the actual statistical distribution of the target spectra-namely the mean. Healy and Slater (1999) have demonstrated an improvement to this strategy. Their algorithm computes a space of allowed target distributions based on variations in illumination and atmospheric effects, and then projects a pixel spectrum onto that space. A classification decision is made based on the angle between the pixel spectrum and the target hyperplane when simple normal statistical models are assumed for the noise. Such an algorithm provides more robust performance, but necessitates exhaustive calculations of potential signatures from each target of interest, and while they reduce the dimensionality of the target hyperplane as much as possible, they do not weight regions of this hyperplane where the target spectra are most likely to lie based on statistical distributions. Furthermore, the target hyperplane is composed by assuming only a mean spectral reflectance, assuming all variations are due to changes in environmental parameters.

This thesis represents an attempt to understand the actual statistical distributions that can be expected from targets in spectral imagery. In this investigation, analysis is restricted to single bands, but it is seen that deviations from normal distributions are common, even when there is enough radiance that one might assume that such a model is

sufficient. Chapter II gives a brief overview of the history of hyperspectral imagery and gives a basis for the work accomplished in this study. Chapter III provides a detailed description of the spectral imagery scene used for this study and the process by which histograms are generated. Several histograms display the nature of the regions of interest that are examined. Chapter IV covers the statistical analysis of the hyperspectral data, including total radiance modeling, single band modeling, and goodness-of-fit testing. Chapter V is a discussion of the implications of the work in this thesis and includes a discussion into further work to be considered in this area of research. Chapter VI concludes the thesis.

THIS PAGE INTENTIONALLY LEFT BLANK

II.   BACKGROUND

A.   HYPERSPECTRAL IMAGING HISTORY

This thesis deals with a specific field of remote sensing called imaging

spectrometry, also known as hyperspectral imaging (HSI). HSI sensors are capable of

providing remotely sensed data in extremely narrow spectral bands, which permit the

identification of materials on the earth's surface. Every material has different reflective

and emissivity characteristics that vary with wavelength and material composition. A

hyperspectral sensor obtains a spectrum of reflected and self-emitted photons that can be

used to uniquely identify a certain material. The concept of imaging spectrometry is

displayed in Figure 2.1 (Vane and Goetz, 1988, p.2). Each picture element or pixel in the

scene contains enough information for the reconstruction of a reflectance spectrum. This

allows the narrow-band spectral features to be used for material identification.
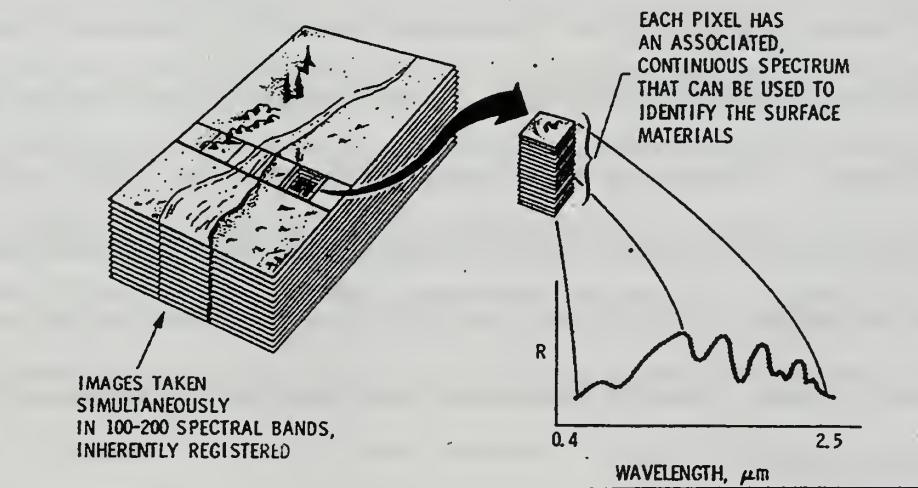
Figure 2.1.Concept for Imaging Spectrometry. From Vane and Goetz, 1988, p.2.

Scientists have been using spectral data for decades to identify numerous characteristics about the earth such as vegetation growth rates, geologic mapping, forest fire damage, shallow water monitoring of oil spills, and mineral identification. Recent advancements in HSI have led to sub-pixel object identification and various military applications such as the ability to recognize a target from its spectral signature without spatially resolving it, ability to discriminate targets from backgrounds, decoys, and each other, and the increased potential for robust automated detection using tractable algorithms, even in highly segmented and cluttered scenes (Kailey and Illing, 1996, p. 15).

The spectral imaging area in remote sensing has undergone tremendous development in the last three decades. In 1972, the Landsat I satellite was launched with one of the first multispectral imaging devices onboard, the Multispectral Scanner (MSS). This sensor used a system of four filters and detectors to split image data into multiple spectral bands in which the detector elements mechanically scanned a swath of the earth's surface perpendicular to the instrument flight path in a sweeping-like motion. Shortly after the development of the MSS, a second-generation Landsat sensor, the Thematic Mapper (TM) was created with seven filter/detector pairs. While both the MSS and TM were significant advances in remote sensing technology, they lacked the spectral resolution and number of spectral bands needed for more precise surface measurements. A. F. Goetz, a pioneer in imaging spectrometry, best described this problem in the following paragraph:

> The need for imaging spectrometry grew out of recognition in the 1970's based on laboratory and field spectral measurements, primarily of rocks and soils, that multispectral imaging in 4 spectral bands with the Landsat MSS was not adequate to discriminate among, much less identify, minerals on the earth's surface that

6

were important in resource exploration and environmental assessment. However, in the 1970's the technologies of optics, detectors, digital electronics and affordable computing, equivalent to the workstation platforms that we are presently accustomed to, did not exist. The ideas and the vision were also too new to gain immediate acceptance by program managers holding the new-technology purse strings (Goetz, 1995, p.3).

The next major advancement in spectral imagery came in 1981 when the Shuttle Multispectral Infrared Radiometer (SMIRR) flew on the second flight of the NASA Space Shuttle. The SMIRR proved that identification of surface mineralogy could be made from orbit with continuous, narrow-band spectral measurements (Goetz, 1995, p.3). SMIRR was a 10-channel radiometer, where three of the channels were only 10 nm wide and centered 10 nm apart around 2.2 μm. These channels, and a fourth centered at 2.35 μm, made it possible to positively identify limestone and the mineral kaolinite (Goetz, 1995, p.3).

The stage was then set for the next generation of remote sensing systems which contained dispersing elements allowing acquisition of spectral data over hundreds of narrow, contiguous spectral bands. These new systems came to be called imaging spectrometers, a name that is still used today. The Airborne Imaging Spectrometer (AIS) was built in the early 1980's as a test bed for higher resolution imaging spectrometers and yielded a number of advances in detector technology and data analysis (Vane and Goetz, 1988). The AIS made use of a grating spectrometer that separated a signal into 128 adjoining bands in the spectral region from 1.2 to 2.4 μm, instead of a system of filters and detectors. In order to be able to cover this spectral region with 9.3 nm sampling, the grating was stepped through 4 positions in the 40 ms needed to traverse one pixel on the ground. AIS 2, a later version of the sensor, incorporated a larger element array and

imaged over the 0.8–2.4 μm region. The AIS system, however, had its problems, among

them, lack of full spectral coverage in the visible and shortwave infrared and also a very

narrow field-of-view (FOV). The success of AIS led NASA to create an imaging

spectrometer program from which the next generation aircraft system, the Airborne

Visible/Infrared Imaging Spectrometer (AVIRIS) was developed.

AVIRIS came into operation in 1987 and is still one of the primary providers of

imaging spectrometer data to the research community today (Vane, 1993). It covers the

0.4-2.4 μm region with 224 spectral bands and has a resolution on the order of 10 nm,

allowing it to detect most absorption features. The AVIRIS optical system consists of a

scanner, four spectrometers, and a calibration source, all coupled by fiber optics. In fact,

AVIRIS was the first spectral imaging system that was radiometrically calibrated, making

it possible to apply atmospheric models to obtain surface reflectance (Goetz, 1995, p.6).

AVIRIS has acquired data for a wide variety of studies including coastal zone water flow,

snow grain size measurement, determination of water status in vegetation, and coniferous

forest damage. Also, the increased spectral range in the visible region provided by

AVIRIS allows biologists to study important reactions in shallow water biology.

The Hyperspectral Digital Imagery Collection Experiment (HYDICE) is one of

the newest, high-resolution airborne imaging spectrometers in use today (Rickard, 1993).

It covers the 0.4-2.5 μm spectral region with a 10.2 average spectral resolution. At an

altitude of 6 km, its spatial resolution is 3m over a 636m swath when flown at low

altitude (5,000 feet). Also, HYDICE uses a 210-element indium antimonide area detector

array which is designed to cover the full spectral range using a single optical path design.

8

This is an impressive simplification of the AVIRIS design, which uses multiple spectrometer/detector pairs and an optical subsystem coupling.

Another system of note is the Midcourse Space Experiment (MSX) satellite, which carries a hyperspectral imaging suite built by John Hopkins University Applied Physics Laboratory called the Ultraviolet and Visible Imagers and Spectrographic Imagers (UVISI). This sensor suite covers the wavelength region .11-.9 μm and provides a spectral resolution on the order of .5-4.3 nm. The next generation spaceborne spectral sensor is the Moderate Resolution Imaging Spectrometer (MODIS) which is on-board the Earth Observing Satellite (EOS-1). This sensor encompasses the earth every 1 to 2 days with a spectral resolution of 10-15 nm in the visible (Hernandez-Baquero, 1997, p.11).

B.   HYPERSPECTRAL DATA ANALYSIS TECHNIQUES

The detection of a certain target or material is achieved by using the large amount of information available from its observed spectrum. Thus, the problem presented is to localize spatially the parts of the spectrum that are characteristic of the target material. There are many different techniques that can be applied to this problem; however only the most prominent will be briefly discussed here.

Stefanou (1997) surmised that there are five major strategies for the analysis of hyperspectral data. Figure 2.2, shows a listing of these five strategies and the key points of each. There are different methods that can be applied within each strategy as well, however, they will not be discussed in this thesis.

```
┌─────────────────────────────────────────────────────────────────────────┐
│ Spectral Imagery Analysis Strategies                                      │
└──┬──────────────────────────────────────────────────────────────────────┘
   │  ┌──────────────────────────────────────────────────────────────────┐
   │  │ Transformation and Projection                                     │
   ├──┤ - Linear transformation performed on each pixel                   │
   │  │ - Goal is to find the "right" set of basis functions and project  │
   │  │   image into "target" subspace                                    │
   │  │ - May act as a preprocessing step prior to classification         │
   │  └──────────────────────────────────────────────────────────────────┘
   │  ┌──────────────────────────────────────────────────────────────────┐
   │  │ Classification                                                    │
   ├──┤ - Pixels assigned to spectral classes based on similar statistics │
   │  │ - Stochastic outlook on data                                      │
   │  │ - Assumes spectrally pure pixels                                  │
   │  └──────────────────────────────────────────────────────────────────┘
   │  ┌──────────────────────────────────────────────────────────────────┐
   │  │ Linear Prediction                                                 │
   ├──┤ - Exploits the spatial and spectral redundancy inherent in        │
   │  │   spectral images                                                 │
   │  │ - Each pixel is viewed as a linear combination of its neighbors   │
   │  │ - Originally applied to data compression                          │
   │  └──────────────────────────────────────────────────────────────────┘
   │  ┌──────────────────────────────────────────────────────────────────┐
   │  │ Optimal Band Selection                                            │
   ├──┤ - Goal is to pick the best spectral bands that will discriminate  │
   │  │   a target                                                        │
   │  │ - Band selection process may be deterministic or stochastic       │
   │  │ - Scene dependent                                                 │
   │  └──────────────────────────────────────────────────────────────────┘
   │  ┌──────────────────────────────────────────────────────────────────┐
   │  │ Multiresolution Analysis                                          │
   └──┤ - Goal is to use different spatial resolution images to form a     │
      │   high resolution composite image                                 │
      │ - Exploits spatial correlation between neighoring pixels          │
      │ - Concepts related to and involving wavelets                      │
      └──────────────────────────────────────────────────────────────────┘
```

Figure 2.2. Listing of Spectral Imagery Analysis Strategies. From Stefanou, 1997, p.5.

The first strategy entails the use of linear transformations and projections where data is visualized as belonging to either a signal subspace or a noise subspace (Stefanou, 1997, p.8). A subspace is a linear algebra term that describes vectors with similar characteristics or properties. The inherit problem of the mixed pixel is assumed in most methods of this strategy, with a mixed pixel being an assumed linear mixture of spectrally pure materials called endmembers. Each pixel is treated independently, since no assumptions are made concerning the spatial arrangement of the pixel vectors. This has led to a search for more effective techniques within the strategy of linear transformations and projections where no knowledge of the background is assumed.

The second strategy in analyzing hyperspectral imagery described by Stefanou (1997, p.9) is a pixel classification approach. Observed pixel spectra are assigned to classes based upon similar statistical criteria. This strategy assumes a stochastic outlook

of the data. The mixed pixel problem is not assumed, and the target spectrum is discriminated based on its membership in a class separate from background pixels.

The third strategy is formulated upon the theories of linear prediction. It makes the assumption that each pixel vector is a linear combination of its neighbors, and seeks to take advantage of this relationship. The postulate is to create a residual image in which there is less redundancy spectrally and spatially. The data from this strategy is viewed statistically and is modeled as such (Stefanou, 1997, p.9).

The fourth strategy is the process of optimal band selection. The goal is to select the best original spectral bands of the hyperspectral image that can be used to discriminate the target. The band selection process is guided by a either a statistical or deterministic view of the target and background spectra. No explicit assumption about the linear mixing model is made (Stefanou, 1997, p.9).

The fifth strategy involves the use of multi-resolution techniques. Concepts such as wavelets are used to pick out different levels of detail from the hyperspectral image. The spatial correlation between neighbors is exploited, though a statistical view is not necessarily required (Stefanou, 1997, p.9).

The goal of this research is to have a significant impact on the first two spectral imagery analysis strategies. The linear projection techniques assume a mean or average spectrum of the target pixels with the addition of noise. However, more realistic statistical distributions describe target pixels better than a mean spectral signature. In this thesis, attempts were made to employ second order statistics (variance) along with non-normal models to improve performance of these spectral imagery analysis techniques, although no attempts were made to improve classifiers.

11

In this thesis, *a priori* knowledge was used about HSI data obtained from Davis Monthan Air Force Base to be able to classify individual pixels as being a part of a certain target or background class. Since the spatial arrangement of the scene was known, pixels were selected by close, visual inspection of objects, and by using individual pixel spectra to determine if a certain pixel was indeed from a given target class. After creating regions of interest based on this technique, a detailed, higher order statistical analysis was then performed on each of the regions of interest, as is discussed in Chapter IV. This thesis utilizes The Environment for Visualizing Images (ENVI) version 3.2, Interactive Data Language (IDL) version 5.2.1, and MATLAB version 5.3.1.29215a, for the analysis of the HSI data from Davis Monthan Air Force Base.

III. HISTOGRAMS

A. HYPERSPECTRAL SCENE: DAVIS MONTHAN AIR FORCE BASE

The hyperspectral data used in this thesis was gathered from a Hyperspectral Digital Imaging Collection Experiment (HYDICE) imagery scene of Davis Monthan Air Force Base (DMAFB) in Arizona. Two sets of HYDICE passes (GSD of 1.5 meters) were taken of DMAFB in June and October 1995. This scene was chosen for three primary reasons. The first reason was for its inherently consistent background characteristics. The predominate background material is sand, which, was easy to separate from targets due to its consistent spectral characteristics. The second reason DMAFB was chosen was for its variability of reference target data. There are hundreds of aircraft, clustered by type, positioned in a designated area of the base – the aircraft boneyard. The type of aircraft include, but are not limited to, the A-6 Intruder, P-3 Orion, C-130 Hercules, B-52 Stratofortress, F-4 Phantom, and AV-8B Harrier. The aircraft are also painted in varying paint schemes such as gray, white, and camouflage. The combination of numerous aircraft and varying paint schemes provides a broad selection of spectral targets for hyperspectral analysis. The third reason Davis Monthan was used was for its ability to provide hundreds of classified pixels for each target class. Having an extremely large group of classified pixels helps to ensure that the statistics compiled are accurate and reliable. Figure 3.1 shows an overhead view of the portion of the October and June over-flights of Davis Monthan AFB analyzed here. Various types of aircraft are identified in the figure, as well as a fence line that separates the airfield and nearby road.
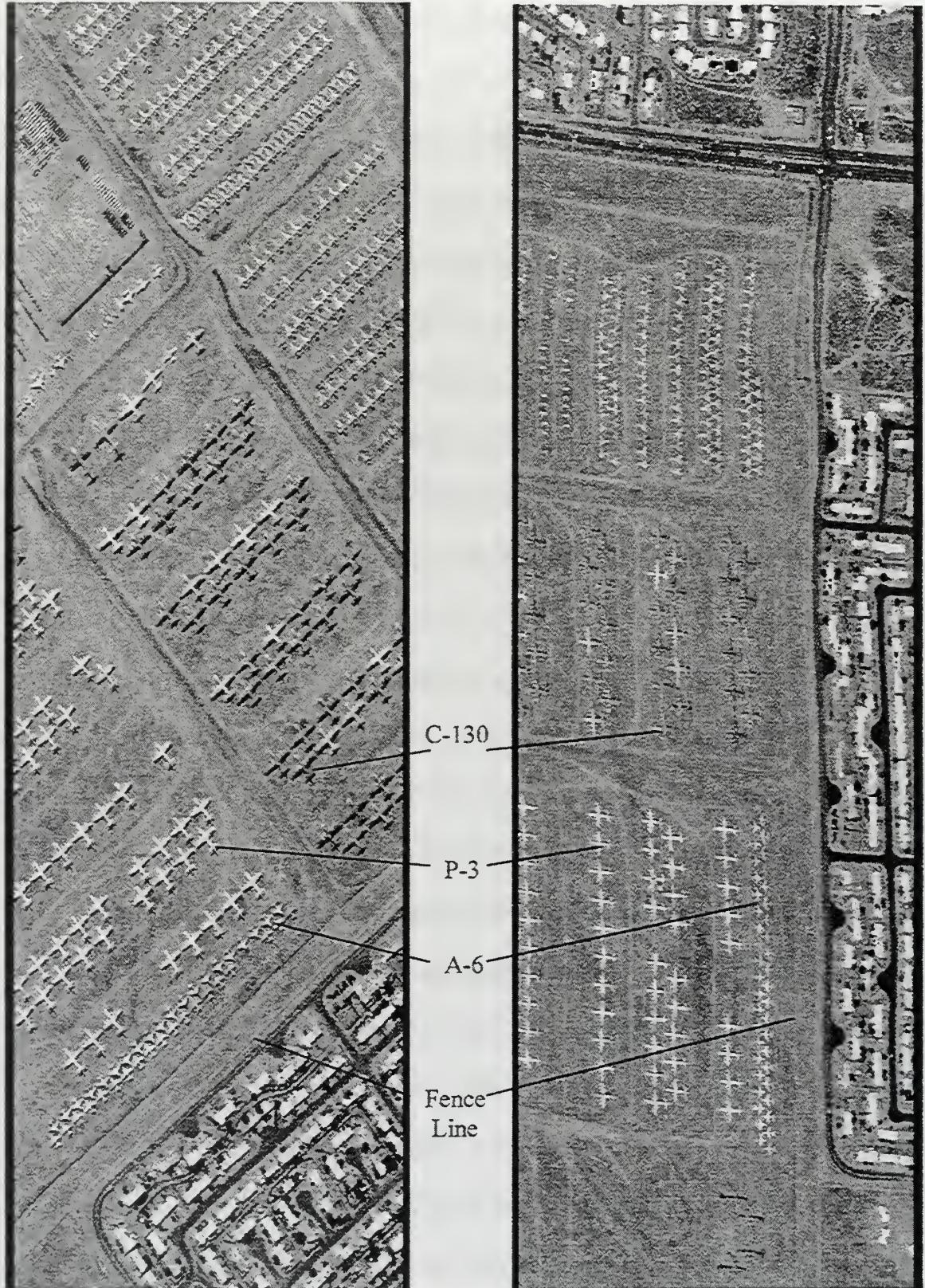
Figure 3.1. HYDICE scenes of Davis Monthan AFB (October scene is to the left).

14

Figure 3.2 shows a portion of area 4, which contains P-3 aircraft, and a portion of area 5, which contains C-130 aircraft. The two areas are separated by a road that runs diagonally across the figure (highlighted in white). The aircraft in both areas are painted in either a white, a gray or a black-brown-green camouflage pattern. Given the fact that areas 4 and 5 have a variety of distinct targets and were covered by both passes of the HYDICE sensor, the majority of the data used in this research was taken specifically from these two areas. Using the data gathered from these two areas enables further study in the difference between passes.
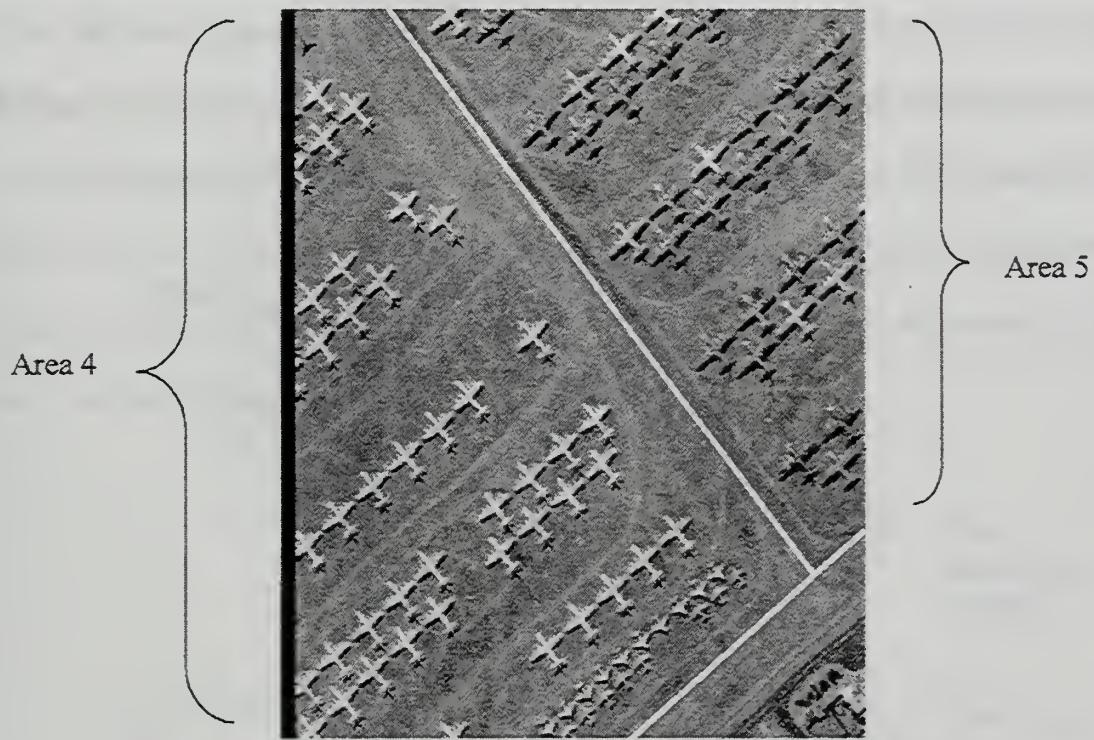
Figure 3.2. Areas 4 and 5 of Davis Monthan scene (October over-flight).

Davis Monthan Air Force Base, given its advantages as a hyperspectral scene, made it easy for regions of interest (ROIs) to be identified. ROIs are groups of individual target pixels that are grouped together to uniquely classify them as being a certain object.

In HSI research, objects are usually classified based on their spectral signature (radiance or reflectance) as a function of wavelength. For the purposes of this study, pixels are grouped into ROIs based on *a priori* knowledge of the types and locations of the aircraft. From these ROIs, histograms were developed that are used for statistical modeling. In this thesis, each ROI was developed through careful analysis of individual pixel profiles in order to establish a common relationship among the pixels, such as being a part of the same P-3 Orion aircraft. This was accomplished in ENVI by zooming into an individual aircraft so that each pixel is clearly discernible. Then, based on pixel profiles and view of the aircraft body, individual pixels were selected and identified as being "pure" and associated with that type of aircraft. Figure 3.3 shows this pixel identification process for a P-3 aircraft. Pixel 1 is clearly on the aircraft, pixel 2 is clearly on the ground, and
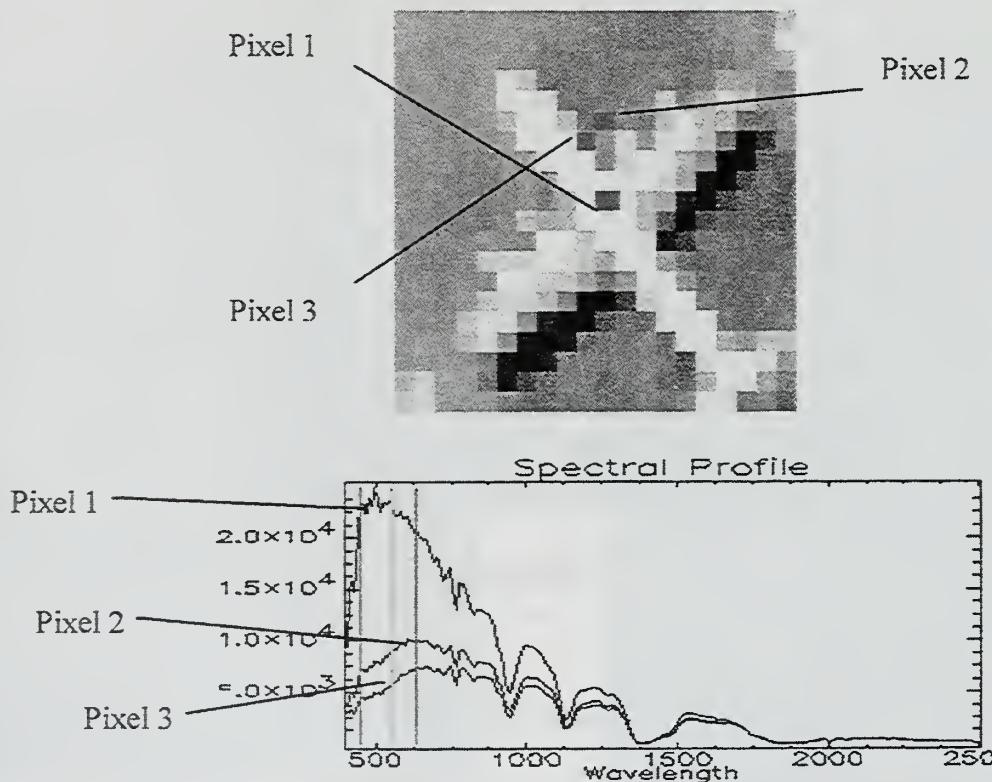
Figure 3.3. Pixel identification process of a P-3.

pixel 3 is a mixed pixel since it falls on the edge of the aircraft and is likely to contain both background and target characteristics. Given that pixels 1 and 3 are pure pixels, the overall shape of the spectrum corresponding to each can be identified and the spectrum of pixel 2, which is intermediate between the two, can be discarded. This enables the selection of only pixels that are reliably on the aircraft in order to classify those pixels as being a part of the P-3 ROI.

For both the C-130 and P-3 ROIs, pixels were selected from dozens of different aircraft that all had the same paint scheme. This prevented ambiguity between pixels since different colors of paint have different spectral signatures and are essentially different spectral targets. Figure 3.4 displays the P-3 aircraft from which pixels were chosen to create the P-3 ROI and the C-130 aircraft that were used to create the C-130 ROI. The selected aircraft are bordered by white lines. Over 1700 "pure" pixels were included in the C-130 ROI and over 1400 in the P-3 ROI to develop a large enough data set to perform reliable statistics.
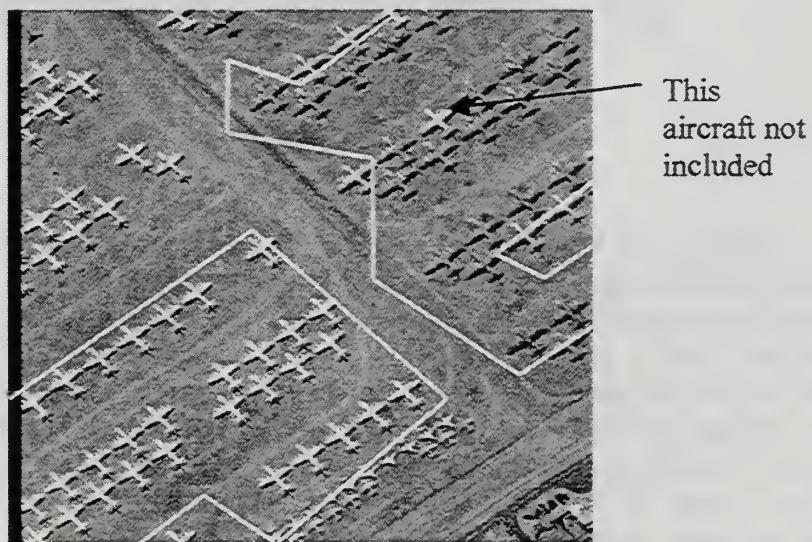
Figure 3.4. P-3 and C-130 aircraft selected for inclusion in Regions of Interests. Only C-130s painted camouflaged were used, here the jagged lines border the selected aircraft.

17

One of the potential applications of HSI is to detect targets that are sub-pixel in size. Thus, a third region on interest, a fence line, was chosen to give statistical information about sub-pixel targets. When a target is smaller than a pixel, the target areas are by definition not pure. The measured reflectance is therefore a combination of the various elements in the spectral scene. Figure 3.5 shows a region of the image that contains the fence line (highlighted in white) that separates the boneyard from the road.

Figure 3.5. Fence line ROI.

Initially, the fence line presented some difficulty in classifying individual pixels; however by comparing one pixel on either side of a fence pixel, a clear indication of the fence line spectral profile became evident. Figure 3.6 shows how a fence line pixel was selected. The top portion is a close-up view of the fence line and the bottom portion shows the pixel spectra.
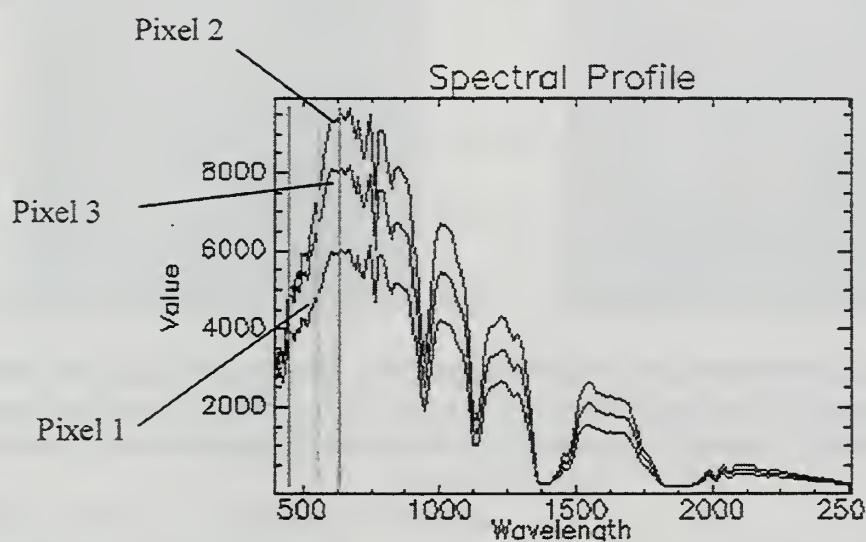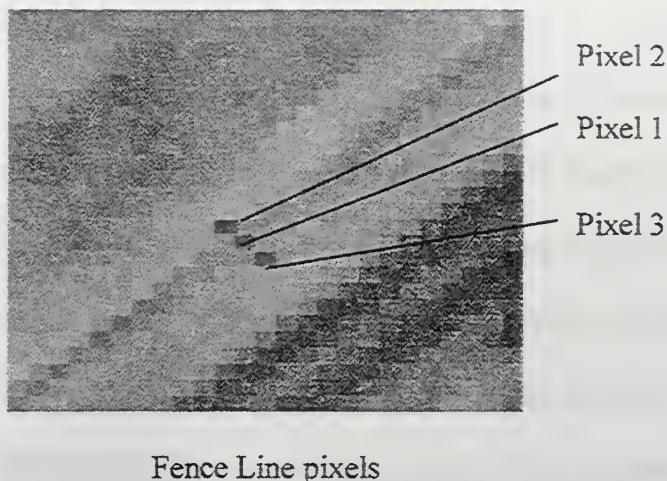
Fence Line pixels

Figure 3.6. Fence line pixel classification.

Pixel 1 is on the top of the fence line with pixels 2 and 3 being on either side of the fence. These three spectral signatures are very similar, but as can be seen from the spectral plots, pixels 2 and 3 show a higher overall radiance value than the profile of pixel 1. This behavior was consistent along the entire fence line and was most likely due to shadowing. Thus, pixels 2 and 3 were eliminated and pixel 1 was selected as being a part of the fence line ROI.

In order to further examine the characteristics of the entire DMAFB scene and to compare potential "targets" with "background", two separate background regions of interest were created. These ROIs are primarily composed of sand and dirt and provide a good basis for comparison with the other selected ROIs since sand and dirt have a higher overall reflectance than the other ROIs. Figure 3.7 shows the areas of the Davis Monthan scene that contain the background ROIs.
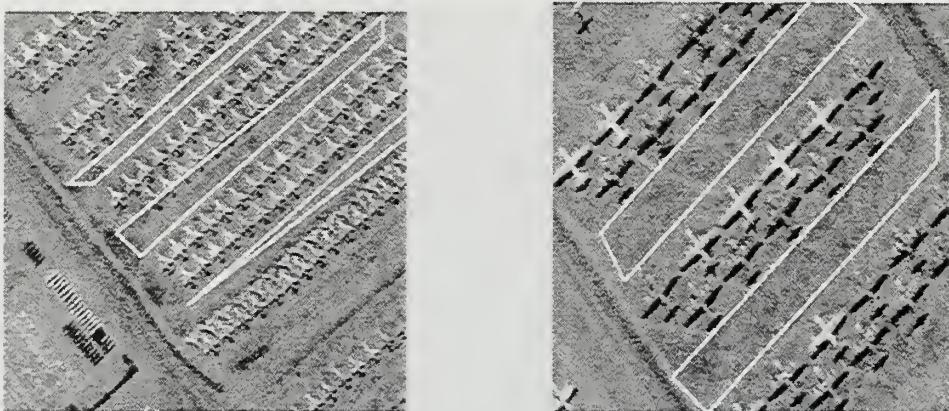
Figure 3.7. Regions of DMAFB for Background ROIs. The picture on the left shows the region from which the first background ROI is taken. The picture on the right shows the region from which the second background ROI is taken. The specific areas are bordered with white lines.

These five regions of interest provided the data from which various histograms where generated and analyzed. Table 3.1 shows the number of pixels included in each ROI. The following section describes the process that created these histograms.

ROI	# of Pixels
C-130	1740
P-3	1689
Fence Line	319
Background 1	5672
Background 2	4725

Table 3.1. Number of Pixels included in each ROI.

B.   HISTOGRAM GENERATION

Three different variables were used for the statistical analysis in this thesis, total pixel radiance, single band normalized and single band unnormalized spectral radiance. The total radiance histograms were used to gain an overall understanding of the data being examined since each histogram exhibits unique characteristics due to illumination variations. A program was written in MATLAB that extracts the total radiance data for each pixel in the regions of interest and creates a histogram of the data. A copy of all the computer programs used in this thesis is included in Appendix A.

1.   Total Radiance Histograms of the Regions of Interest

The histograms presented here represent the total radiance across all bands for each respective ROI. Each histogram exhibits unique characteristics due to illumination variations resulting from spatially changing atmospheric effects, shading, shape, and other physical properties of the material being investigated (Kruse, 1988). The total radiance at each pixel is simply computed as

$$L_{tot} = \sum_{\lambda} L(x, y, \lambda)$$

(3.1)

where $L(x, y, \lambda)$ represents the radiance of each pixel at a certain wavelength (λ). The histogram for the total radiance within the C-130 ROI for the October over-flight is shown in Figure 3.8.

Total Radiance Histogram of C-130 ROI
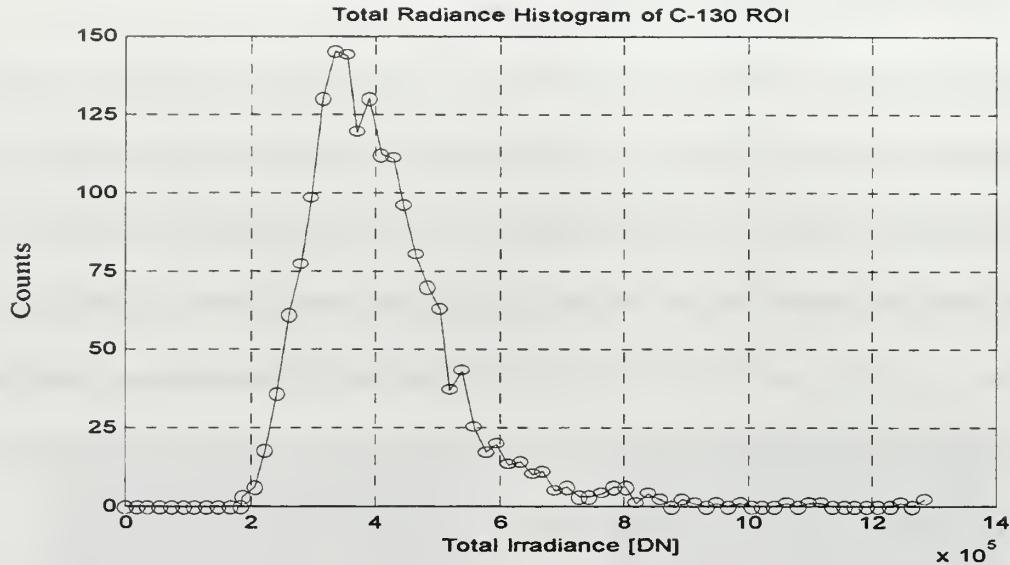
Figure 3.8. Total Radiance Histogram of C-130 ROI.

A purely cursory examination of this histogram reveals that it is not normal in nature but rather skewed to the left with a long right tail. This means that a normal distribution probably will not fit this data well. In the following chapter, different statistical distributions will be fitted to this histogram in an attempt to model it effectively.

Figure 3.9 displays the P-3 ROI histogram. This has a bi-modal distribution, shown as two humps in the histogram, which makes it difficult to model. Further investigation revealed that the cause of the bi-modal distribution was due to fluctuations in radiance along the fuselage of the P-3 aircraft. When selecting pixels for inclusion in the P-3 region of interest, pixels were taken from across the entire surface of the aircraft, including the fuselage and the wings.
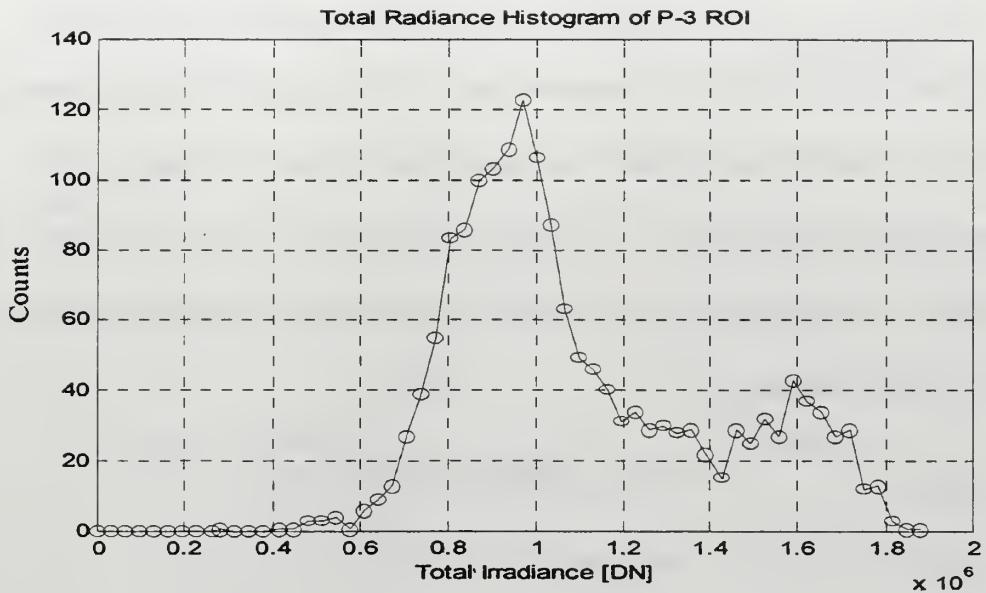
Figure 3.9. Total Radiance Histogram of P-3 ROI.

Figure 3.10 shows histograms of the total radiance across the P-3 fuselage and the P-3 wings. The wings show a single mode distribution.
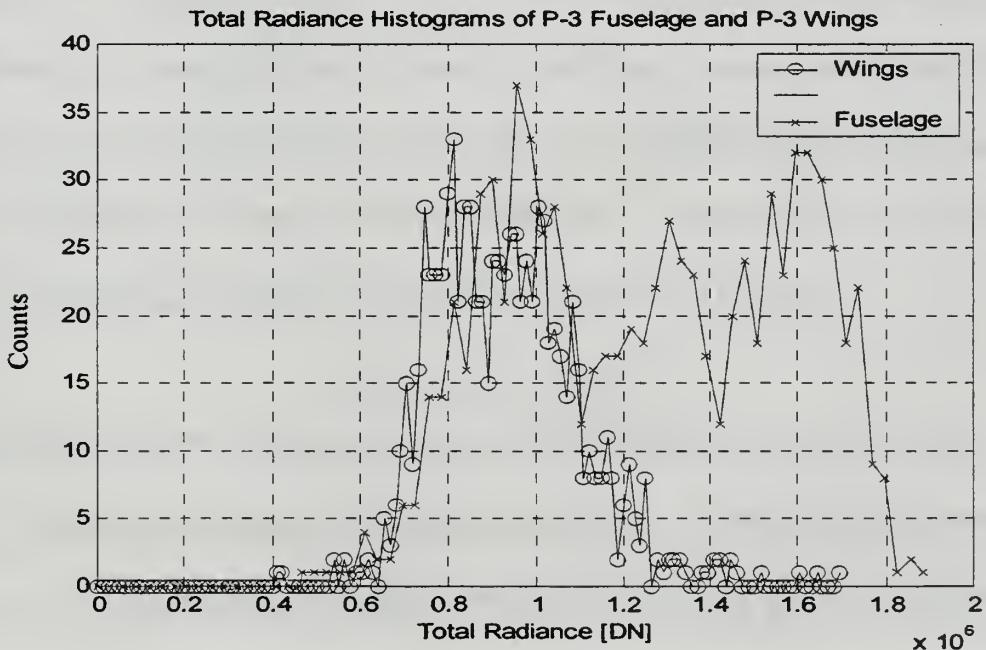
Figure 3.10. Total Radiance Histograms of P-3 Fuselage and P-3 Wings.

23

The fuselage, however, seems to be the contributing factor to the bi-modal nature of the entire P-3 ROI. This feature of the P-3 ROI makes it difficult to adequately model with the statistical distributions considered here. Figure 3.11 displays the histogram for the fence line ROI. This region of interest was chosen in an effort to show the difficulties in analyzing sub-pixel targets. The histogram is slightly skewed to the right with a short, left tail.
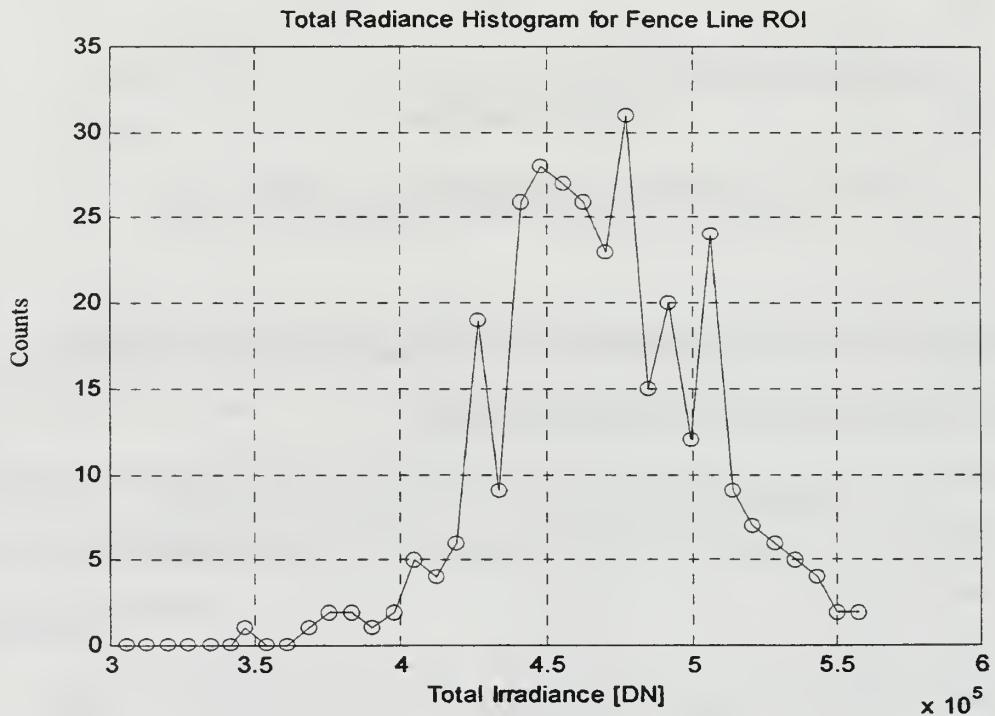
Figure 3.11. Total Radiance Histogram of Fence Line ROI.

Figure 3.12 shows the histograms of the background ROIs. They both are slightly right skewed with short left tails. These histograms will be compared to the same statistical distributions as the previous histograms to check for goodness of fit.

24

Total Radiance Histograms of Background ROIs
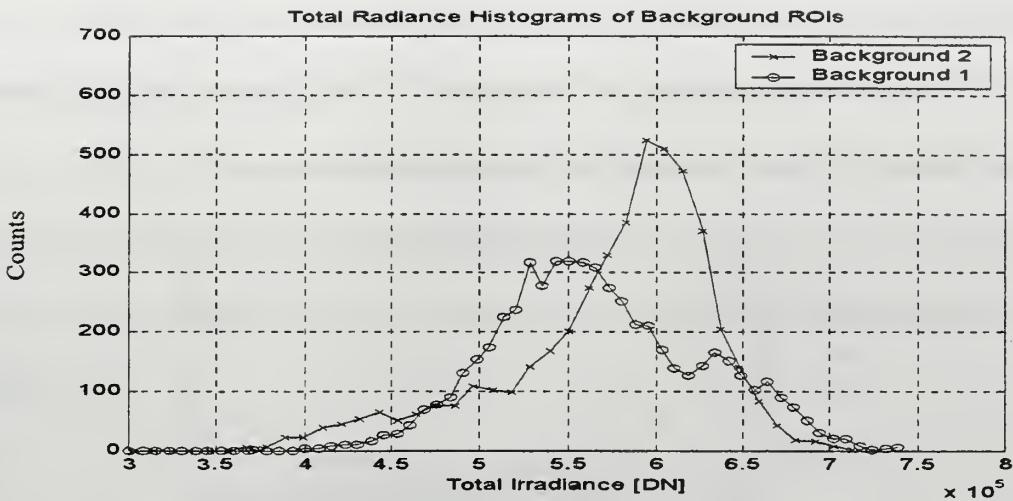
Figure 3.12. Total Radiance Histograms of Background ROIs.

2.   Single Band Histograms of the Regions of Interest

Using IDL, a program was written that generates single band, normalized and unnormalized radiance histograms from selected ROIs. The program first accepts a ROI as input and normalizes the data in that ROI. This computation is a standard operation performed to eliminate the effects of albedo in processing spectral data (Kruse, 1988). A region of interest is normalized by dividing the spectral radiance value of each pixel in the ROI by the total radiance of all bands in that pixel. The normalization function can be written as:

$$\bar{L}(x,y,\lambda) = \frac{L(x,y,\lambda)}{\sum_{\lambda} L(x,y,\lambda)} \qquad (3.2)$$

Equation 3.2 scales the spectrum of each pixel to approximately the same overall relative brightness. The normalization process is critical in reducing the number of ambiguities in spectral data. Much of the variability in a spectral scene is due to apparent illumination

25

differences that arise from shading and curvature of target surfaces. For example,

Figure 3.13 shows two pixel spectral profiles from the P-3 ROI that are not normalized.

Figure 3.14 shows these same spectra after the normalization process.
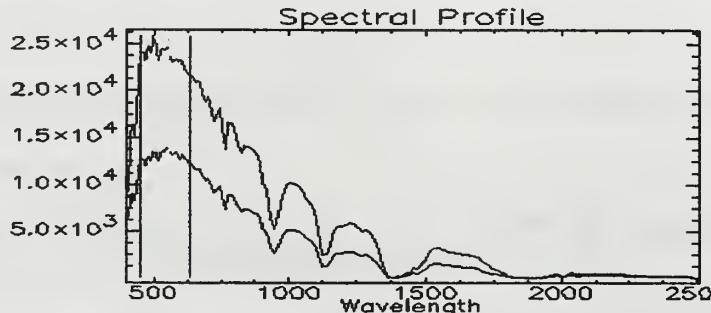
Figure 3.13. Individual P-3 spectra before normalization
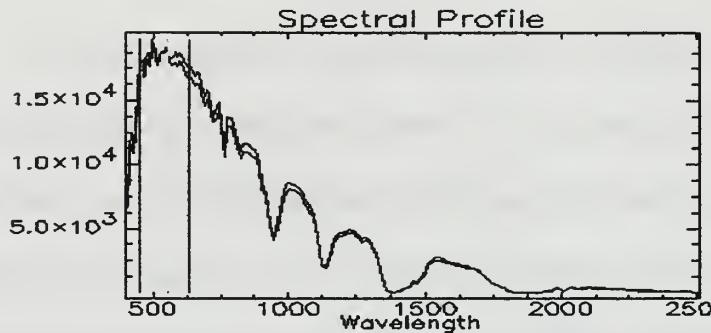
Figure 3.14. Same spectra after the normalization process.

The two bands have been normalized to the same relative brightness, which makes it

more apparent that the two spectra are similar in nature.

After the normalization is complete, a subset of the 210 HYDICE bands was

chosen by investigating the mean spectral signature of the various ROIs. Each ROI has a

mean specific signature due to global factors of illumination, absorption, and reflectance.

Likewise, each pixel spectrum has numerous fluctuations about the mean due to

individual pixel characteristics such as orientation, shading, and imperfections on the

26

target surface. As an example, Figure 3.15 shows a profile of a pixel taken from the fence line ROI. There are peaks and valleys in the spectrum that correspond to different characteristics of the scene, illumination, and atmosphere. In order to understand the statistics of different portions of the spectral signature, bands 50, 68, 75, 80, 88, 94, 114, and 117 were chosen. The points corresponding to these bands are highlighted in Figure 3.15.
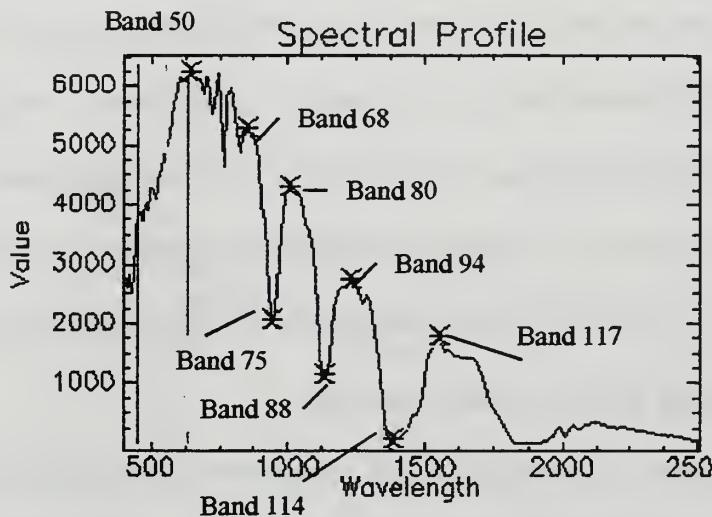

Figure 3.15. Pixel spectrum from fence line ROI with highlighted bands.

Once the bands were selected, the histograms could then be calculated. For each band, the range of normalized radiance was separated into bins for proper viewing of the histogram. The number of bins chosen depends on the ability to give reasonably smooth histograms for analysis. A copy of the program written to compute these histograms is given in Appendix A. The following section will provide examples and discussion of the various histograms that were generated for each region of interest.

The next step in the analysis of the Davis Monthan data is to compute and display the single band histograms for each of the respective ROIs from the October data. Each of the histograms in this subsection is computed using unnormalized spectral radiance. As was described earlier, each ROI is comprised of data that is divided into 210 spectral bands where each band exhibits its own unique spectral signature. By researching the characteristics of several of the spectral bands within a ROI, one can determine which statistical distribution fits best for specific bands, and hopefully gain insight into its overall distribution. The first step is to look at the unnormalized single band histograms and observe the significant differences between the normalized and unnormalized data. The C-130 and P-3 wing ROIs will be presented here as examples of what was seen across all five regions of interest.

Figure 3.16 shows the C-130 unnormalized single band histograms. All of the histograms appear left skewed with right tails.
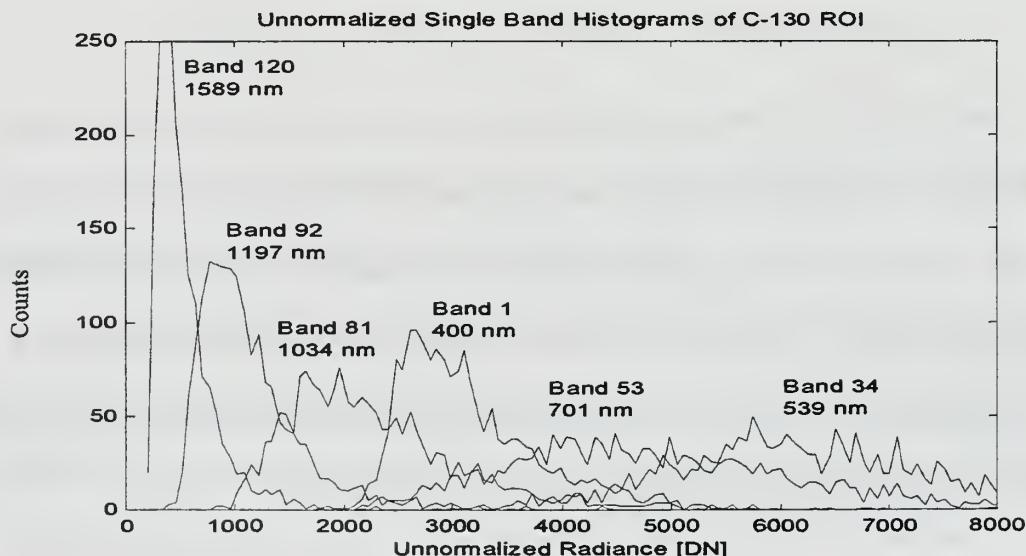

Figure 3.16. Unnormalized Single Band Histograms of C-130 ROI

28

The next set of unnormalized single band histograms are taken from the P-3 wing ROI. Figure 3.17 displays the unnormalized single band histograms taken from the P-3 wing ROI. The histograms of Bands 59 and 84 were not shown in this figure for purposes of clarity since they overlapped the other histograms. Again, the histograms are left skewed with right tails.
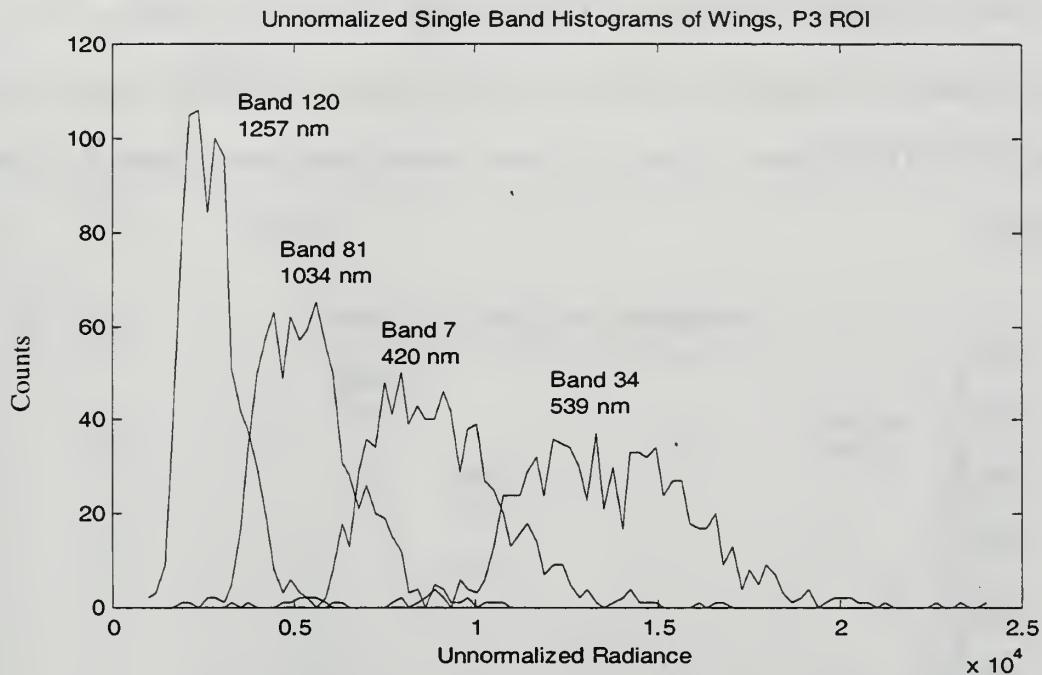

Figure 3.17. Unnormalized Single Band Histograms of P-3 Wing ROI.

The general trend that was seen in Figures 3.16 and 3.17 of left skewed histograms with right tails, was seen in the other regions of interest as well. The following subsection will display normalized radiance histograms from the October over-flight data.

b. Normalized Single Band Histograms of the ROIs (October)

The next step in the analysis of the Davis Monthan data is to compute and

display the normalized radiance single band histograms for each of the respective ROIs

from the October data. Each of the histograms is computed using normalized spectral

radiance, using Equation 3.2. For each of the regions of interest, several histograms of

single bands are displayed from the 210 HYDICE bands. In a separate figure, a few of

these histograms are displayed in an effort to capture the general variability that is seen

across all 210 HYDICE bands. Figure 3.18 shows several histograms of bands from the

C-130 ROI.
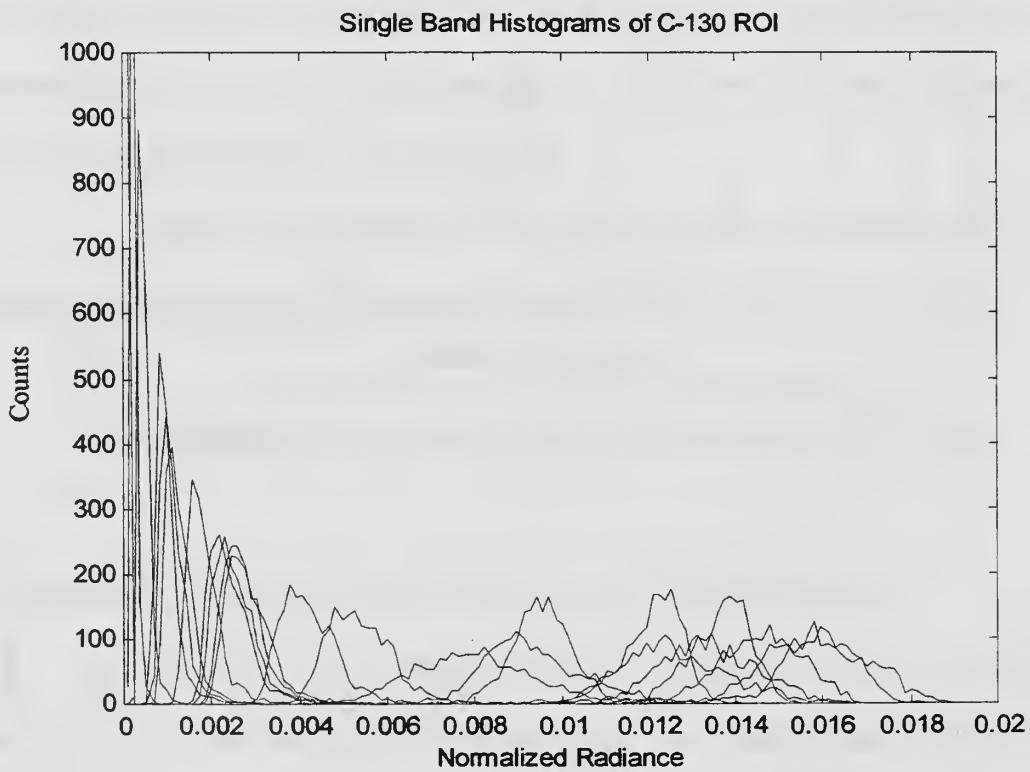
Figure 3.18. Single band histograms of C-130 ROI.

The 22 histograms of individual bands shown in Figure 3.18 show a representative sample of the 210 bands in this ROI. Some of the histograms look symmetric while others are skewed either to the left or right. In order to see the general variability that is seen across all bands, six of the histograms were plotted in Figure 3.19.

Selected Single Band Histograms of C-130 ROI
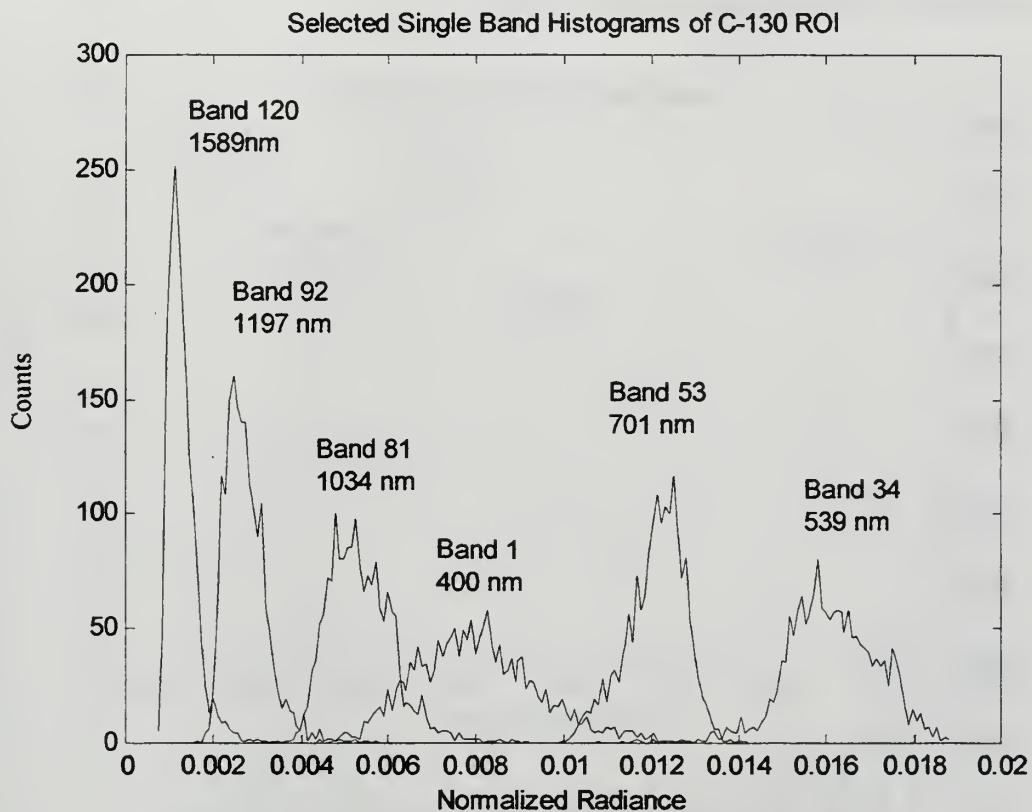
Figure 3.19. Selected single band histograms of C-130 ROI.

These six histograms capture the general variability that is seen across all 210 bands. Bands 1 and 34 are from the visible portion of the spectrum where the measured radiance is high, and while these are the most symmetric distributions, they still appear somewhat skewed. Band 53 is at the edge of the visible, and has its long tail to the left. Bands 81, 92, and 120 are moving through the Near Infrared (NIR) where total response is

decreasing, and the distributions are becoming more skewed to lower values with long right tails. The nature of the normalized single band histograms seen in Figure 3.19 is significantly different from the unnormalized histograms seen in Figure 3.16 where all of the histograms were left skewed with right tails. Figure 3.20 displays various normalized radiance single band histograms from the P-3 ROI.
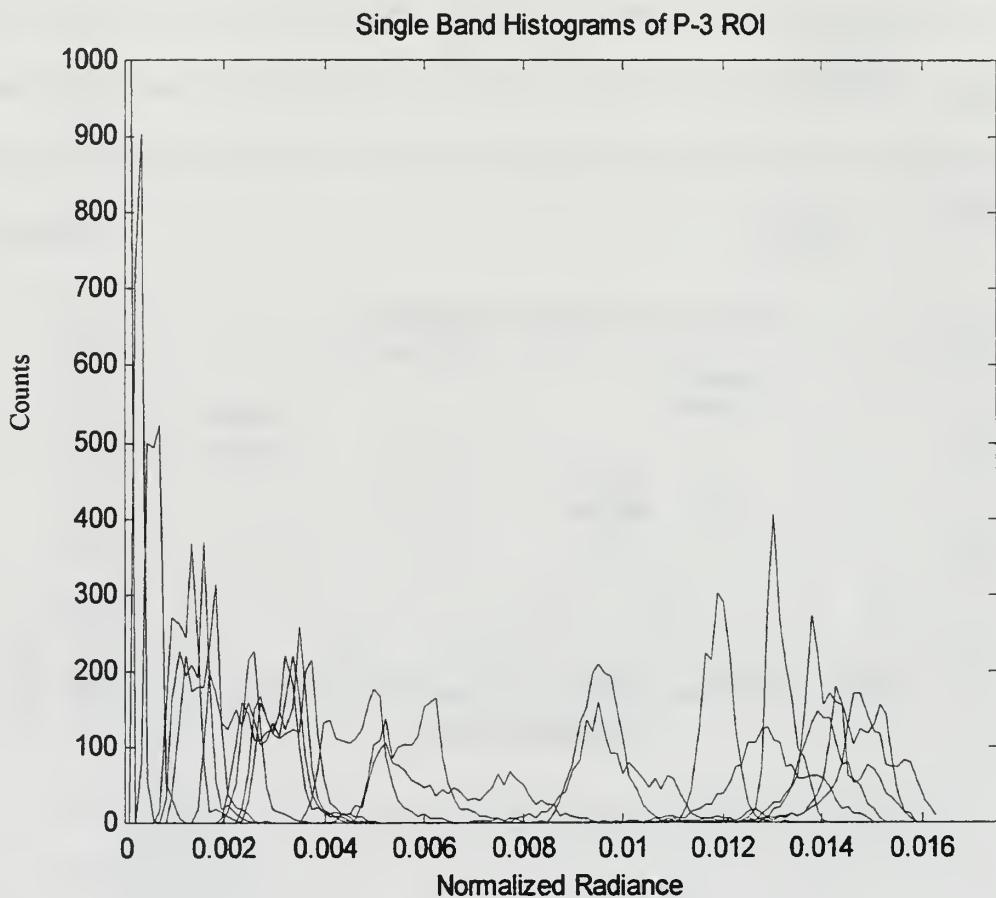

Figure 3.20. Various normalized single band histograms of P-3 ROI.

This figure shows the variability seen across all bands of the P-3 ROI. In some of the histograms, there seems to be evidence of the bi-modal distribution that was seen in the total radiance histograms. By viewing a selection of these single band

32

histograms, the typical nature of the distributions in this ROI should become evident.

Figure 3.21 displays six of the histograms from the P-3 ROI.
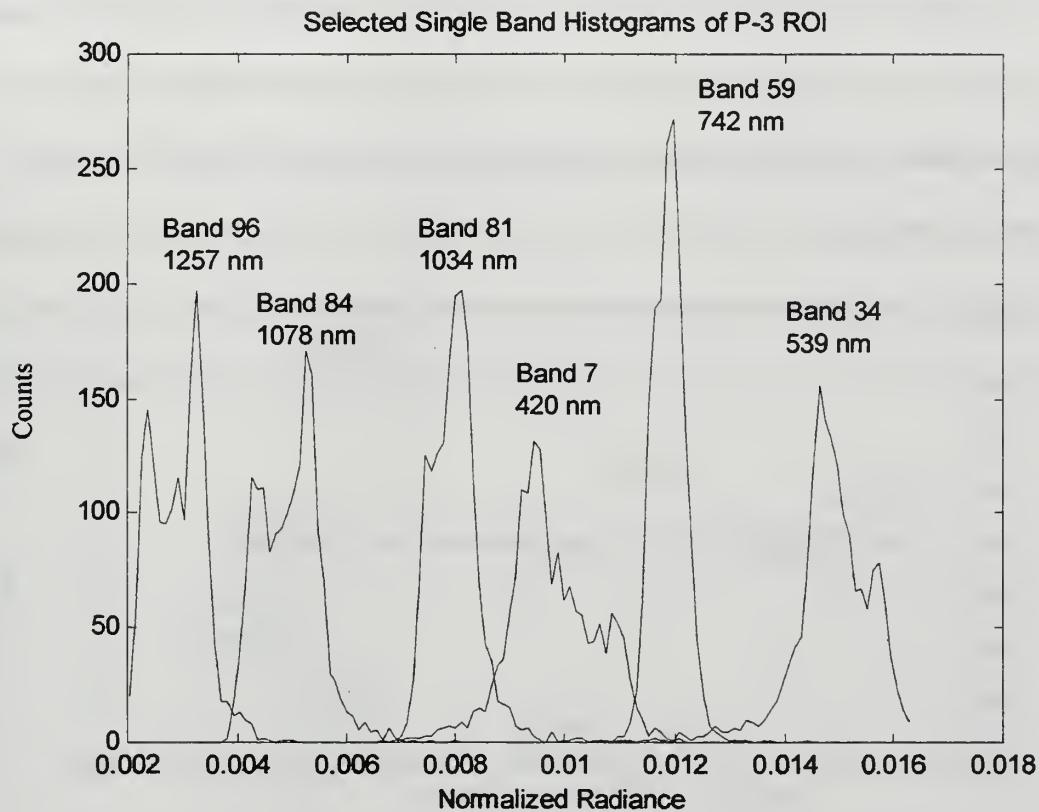
Figure 3.21. Selected Single Band Histograms of P-3 ROI.

The six histograms shown seem to show the bi-modal characteristic of the P-3 ROI. A comparison between these histograms and distributions from just the wing data of the P-3 ROI are necessary to get more favorable histograms for statistical distribution fitting. The following figure shows selected single band histograms from the wing data of the P-3 ROI. The same bands that were used for analysis from the entire P-3 ROI were used in an effort to show the direct comparison between the two sets of data.

33

Figure 3.22 displays six histograms from this data. From the figure, it seems that the histograms, are more unimodal in appearance than the histograms from the entire ROI. This could be an indication, as was discussed previously, that the fuselage data is skewing the results of the P-3 ROI and causing a bi-modal appearance of the histograms. Also, the histograms below are much different from the unnormalized P-3 wing histograms which were all left skewed with right tails. Figure 3.23 shows six selected histograms from the fence line ROI.
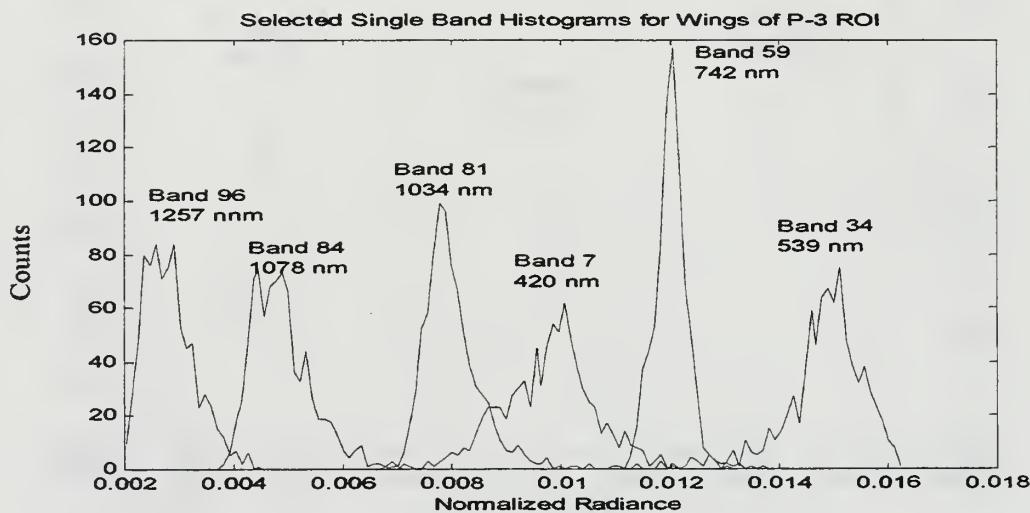

Figure 3.22. Selected single band histograms for wings of P-3 ROI.
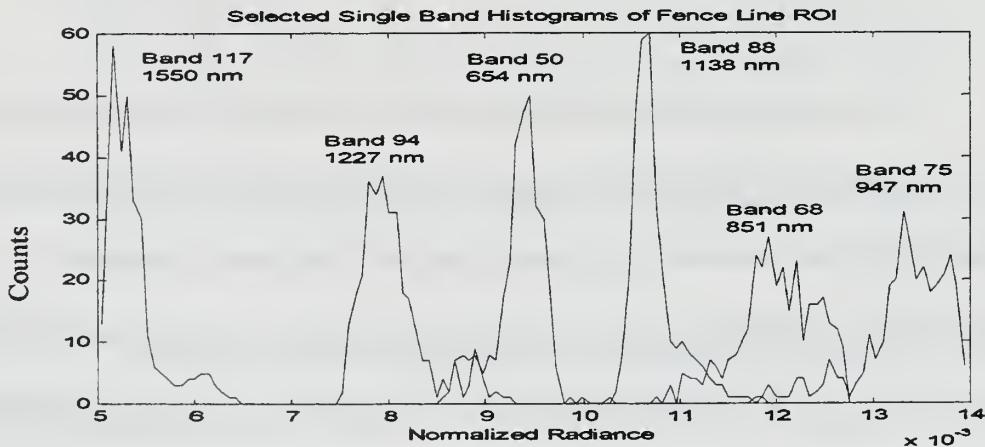
Figure 3.23. Selected Single Band Histograms from Fence Line ROI.

34

In Figure 3.23, bands 50, 68, and 75 all appear right skewed with left tails and bands 88, 94, and 117 all appear left skewed with right tails. Bands 80 and 114 from Figure 3.15 are not displayed here since their histograms overlapped the histograms of Bands 68 and 117 respectively. However, the general nature of the fence line spectrum is still captured using the above six bands.

The next figure displays selected single band histograms from the first background region of interest. Since the background is mostly comprised of sand, it exhibits a very even, spectral distribution with few irregularities. Figure 3.24 displays the six histograms that most closely correspond to the nature of all 210 bands across this ROI.
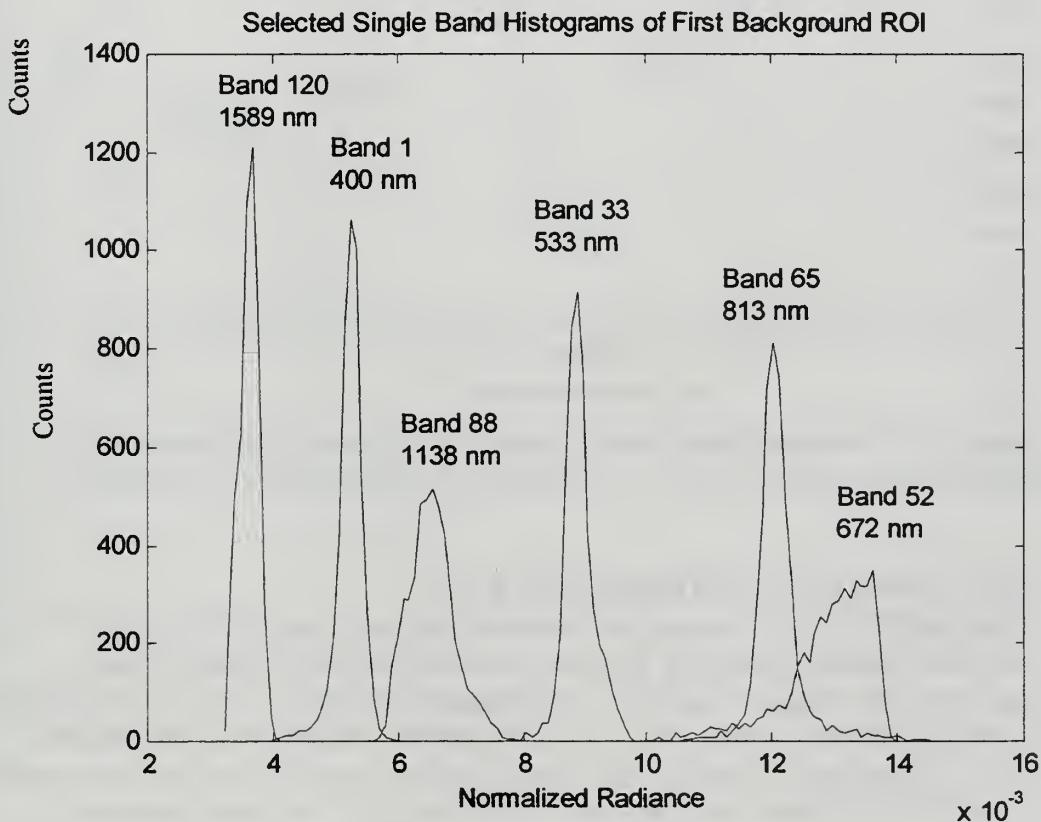

Figure 3.24. Selected Single Band Histograms of First Background ROI.

The histograms in Figure 3.24 are fairly symmetric, with the exception of Band 52, which is skewed to the right. The following chapter will discuss the ability of the normal distribution to fit these single band histograms. As a comparison, Figure 3.25 displays single band histograms from the second background ROI. This ROI was comprised of sand and dirt dominated pixels that were located a few hundreds of meters away from the first background ROI. The similarity in shape of the single band histograms of the second background ROI with the single band histograms of the first background ROI can be seen.
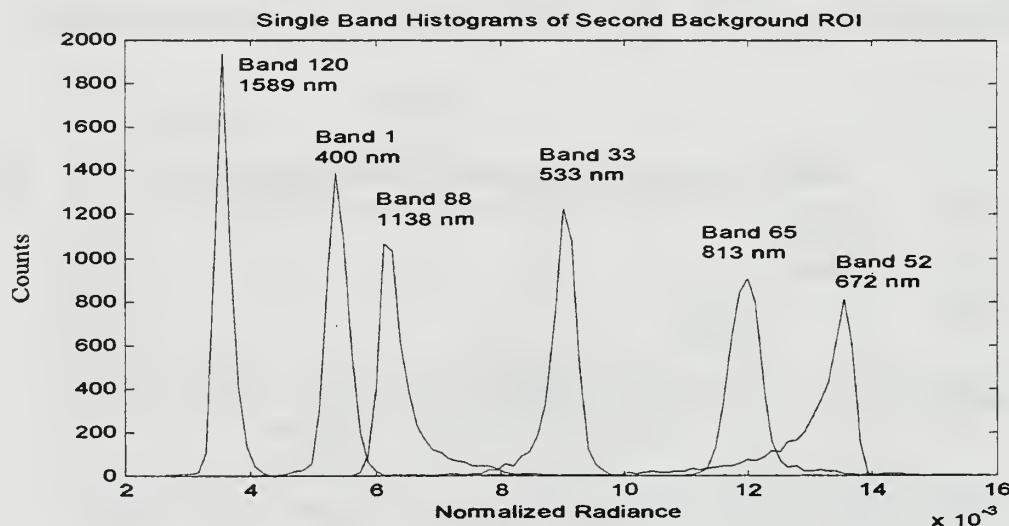

Figure 3.25. Selected Single Band Histograms of Second Background ROI

C.   JUNE 1995 HYDICE OVERFLIGHT DATA

The Davis Monthan data that has been presented thus far was taken from a HYDICE image taken in October 1995. This next section will describe data that was taken from a HYDICE image from the June 1995 over-flight. The three regions of interest created from the June data, a C-130, P-3 Wings, and P-3 Fuselage ROI, were

made using essentially the same aircraft as the October data. As was done earlier in this chapter, the total radiance histograms of each ROI will be examined first, followed by the single band normalized and unnormalized histograms.

Figure 3.26 displays the total radiance histogram of the C-130 ROI. It appears slightly left skewed with a right tail, which is very similar to the October C-130 total radiance histogram. The fact that the histograms look the same is important and will be discussed in greater detail in Chapter V.
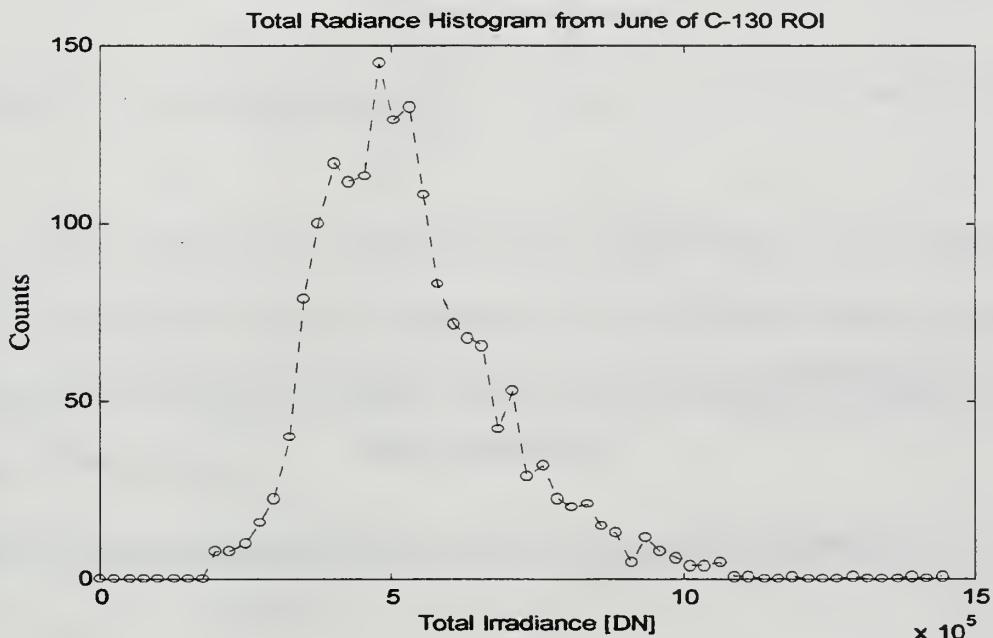

Figure 3.26. Total Radiance Histogram of C-130 ROI from June Over-flight

Figure 3.27 displays the total radiance histograms of the P-3 Wings and P-3 Fuselage ROIs. Again, these two histograms are similar in shape to the total radiance histograms from the October over-flight data. The bi-modal characteristic of the P-3 data is clearly evident in the histogram of the fuselage.
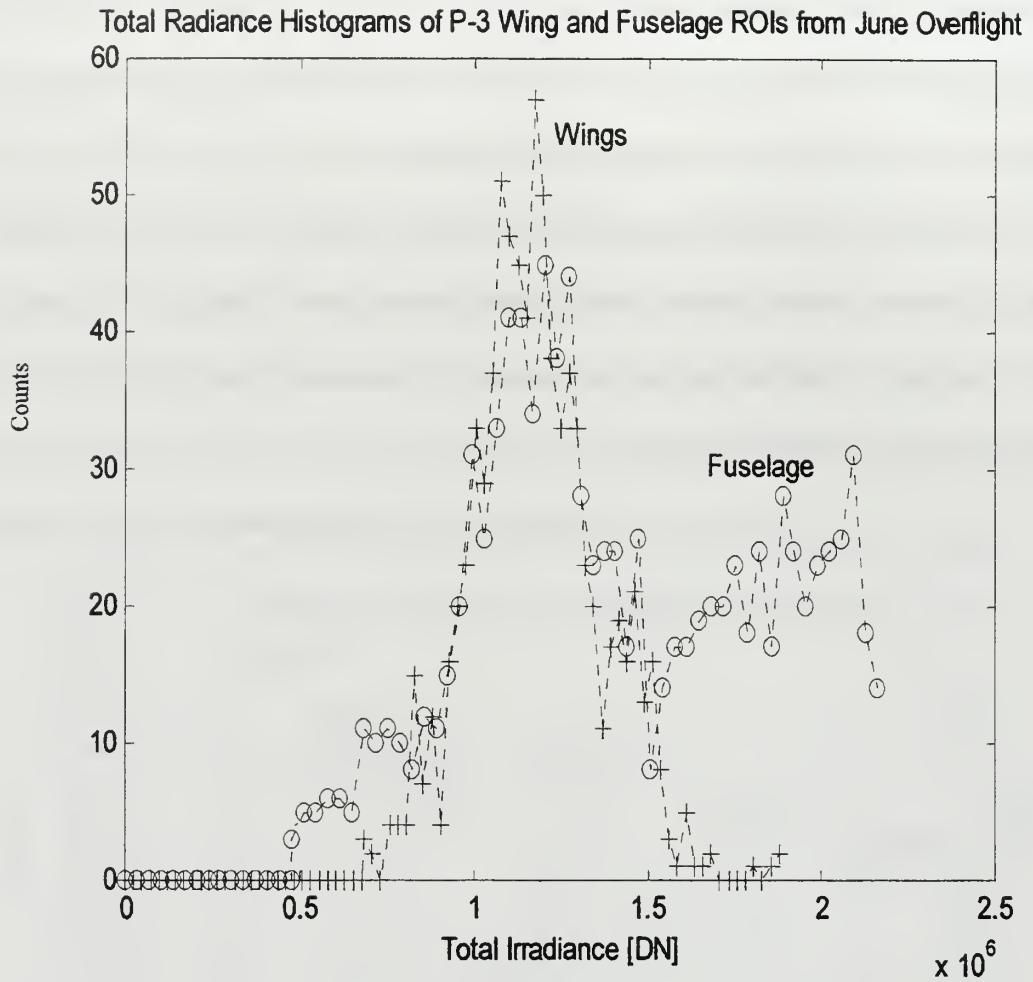
Figure 3.27. Total Radiance Histograms of P-3 Wings and Fuselage from June Overflight

The next histograms that will be displayed are the unnormalized single band histograms from the June over-flight data. Figure 3.28 shows the unnormalized radiance single band histograms of the C-130 ROI. The appearance of these histograms is nearly identical to the unnormalized single band histograms from the October data that were shown in Figure 3.16.

38
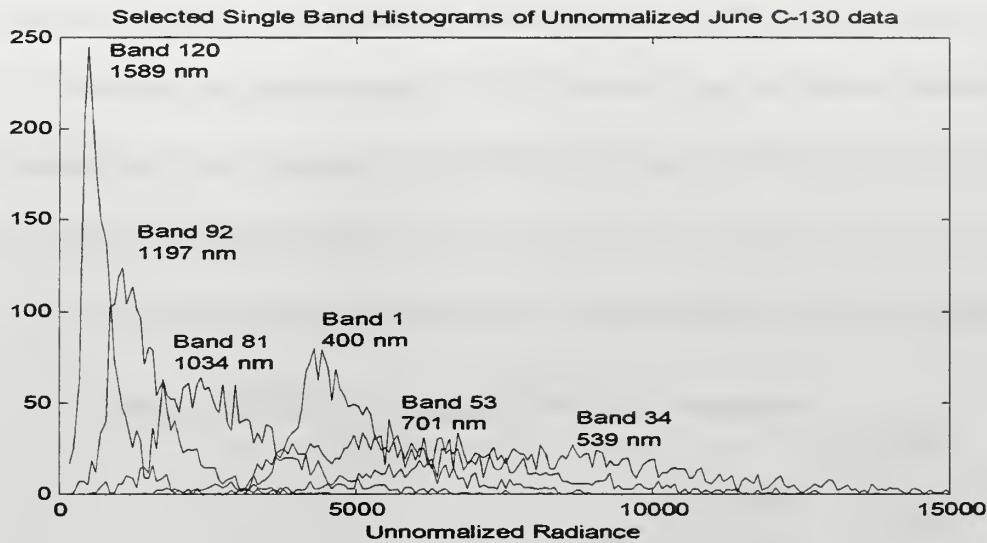
Figure 3.28. Unnormalized Single Band Histograms of C-130 from June Over-flight

Figure 3.29 shows the unnormalized single band histograms of the P-3 Wings ROI. These histograms have a like appearance to the unnormalized histograms displayed in Figure 3.17 and based on the results from the statistical analysis in Chapter IV, exhibit similar distribution fits.
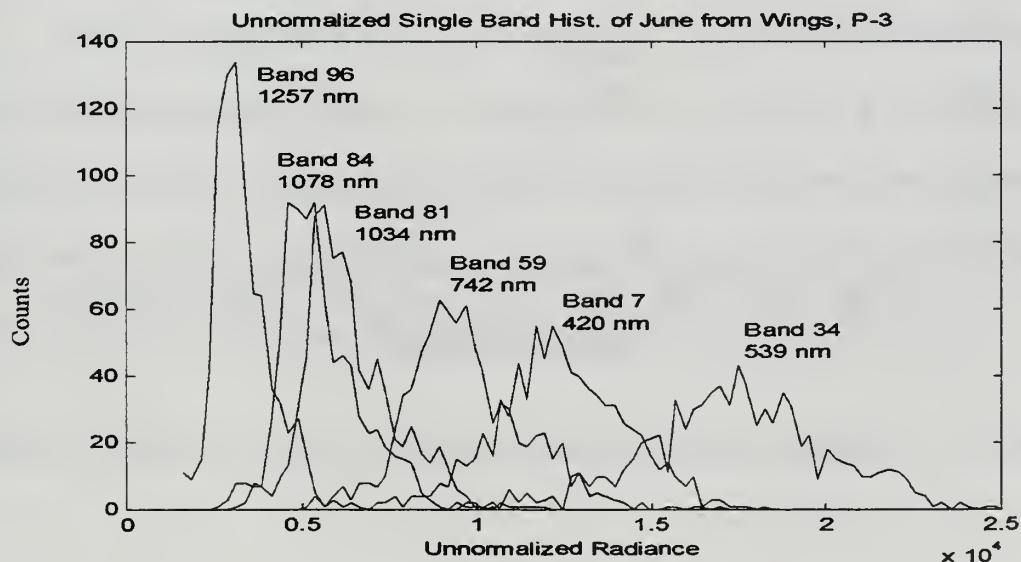
Figure 3.29. Unnormalized Single Band Histograms of P-3 Wings from June Over-flight

The next figures shown are the normalized single band histograms from the ROIs of the June over-flight data. The normalized single band histograms for the C-130 ROI are shown in Figure 3.30. They are very similar in appearance to the normalized histograms of the October C-130 ROI that were shown in Figure 3.19, which is not surprising given the fact that the total radiance histograms were similar as well.
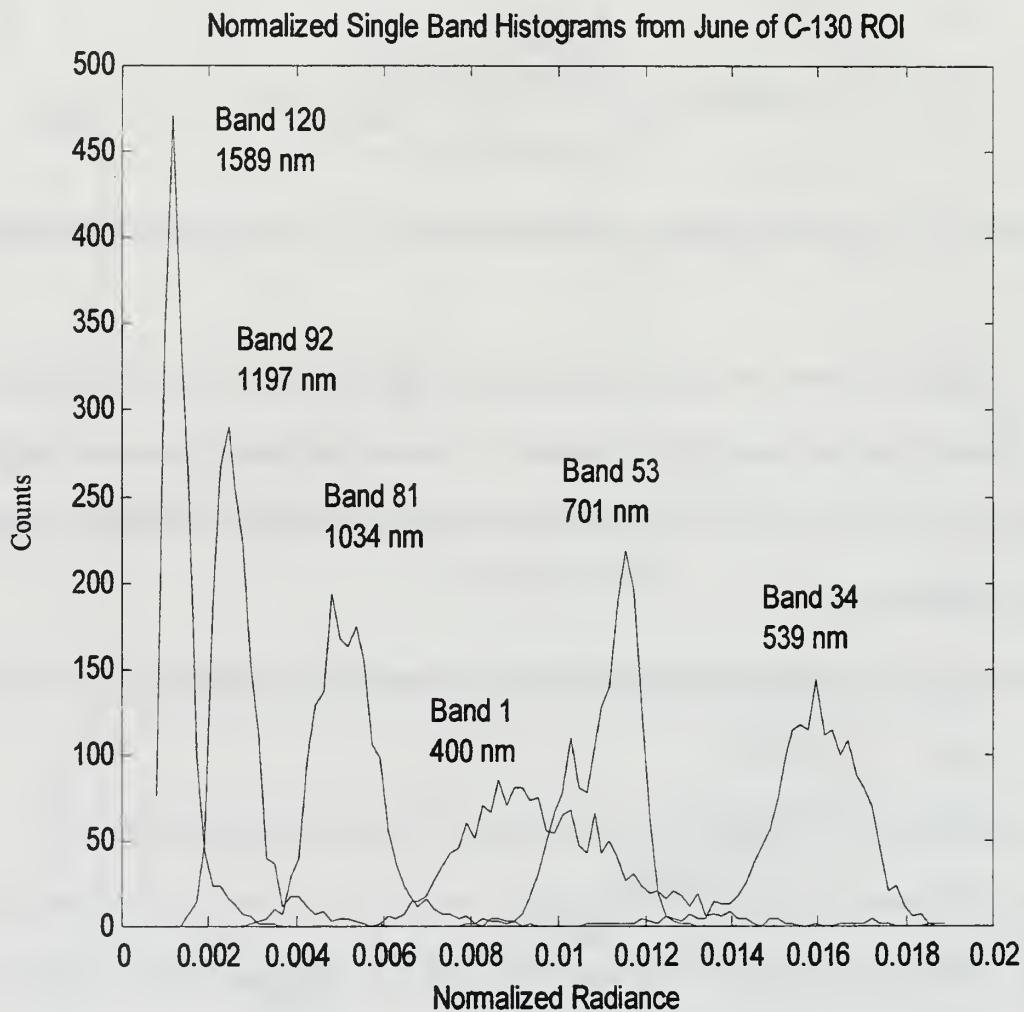

Figure 3.30. Normalized Single Band Histograms of C-130 ROI from June Over-flight

Figure 3.31 displays the normalized single band histograms from the P-3 Wings ROI. The June data histograms are slightly different in appearance from the October data histograms, but each band histogram from the June data is still skewed the same way as each of the band histograms in Figure 3.22.
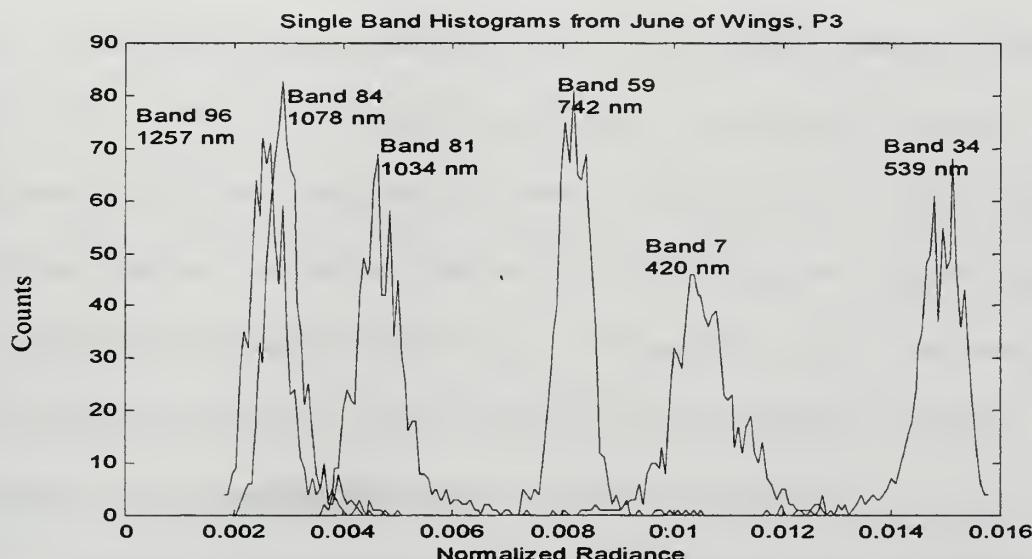
Figure 3.31. Normalized Single Band Histograms of P-3 Wings from June Over-flight

This chapter described the Davis Monthan Air Force Base hyperspectral scene and the different aspects of the spectral data that was extracted from it. Total pixel radiance, unnormalized single band, and normalized single band histograms were displayed for various regions of interest from both the October and June over-flights in an effort to show the wide range of variability that exists in hyperspectral data.

THIS PAGE INTENTIONALLY LEFT BLANK

IV.   STATISTICAL MODELING

A.   BACKGROUND OF STATISTICS USED

In this thesis, several statistical distributions were fitted to the Davis Monthan data. The maximum likelihood estimators (MLEs) of the gamma, normal, and lognormal distributions, as well as the Weibull in some instances, were compared to the total radiance and single band histograms of the selected regions of interest. The χ^2 goodness-of-fit test, a test that computes how well the MLEs fit the histograms, was used to provide a numerical basis for analysis of the MLE fits.   This section will give a brief overview of the statistical distributions, maximum likelihood estimators (MLEs), and the χ^2 goodness-of-fit test used for this research.

The method of maximum likelihood estimation is used to find the parameters of a given distribution that have the best chance of producing the observed data. The likelihood function, gives the probability of realizing the observed sample as a function of the possible parameter values, the mean (μ) and variance (σ^2) for the normal and lognormal, and α and β for the gamma and Weibull. Maximizing the likelihood gives the parameter values that are most likely to have produced the observed sample (Devore, 1995, p.266). The method of MLE was used in this thesis to fit parameters to several different histograms of the total radiance as well as the single band normalized radiance within a respective region of interest.

Based on the shapes of the histograms observed in Chapter III, the normal, Weibull, gamma, and lognormal probability distribution functions were chosen in an

attempt to approximate the functional shape of the actual histograms for the DM data. The likelihood function for a normal probability distribution is

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)},$$ (4.1)

where μ is the mean and σ^2 is the variance. To find the maximizing values of μ and σ^2, we must take the partial derivatives of $\ln(f)$ with respect to μ and σ^2, equate them to zero, and solve the resulting two equations. Normal pdfs are always symmetric and can produce both positive and negative observations. The likelihood function for the lognormal distribution is

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2 x}} e^{-[\ln(x)-\mu]^2/(2\sigma^2)},$$ (4.2)

where a nonnegative random variable x is said to have a lognormal distribution if the random variable $y = \ln(x)$ has a normal distribution. Note that in this case, μ and σ^2 are not the mean and variance of x, but of $\ln(x)$. Lognormal pdfs are always skewed left with long right tails. The Weibull and gamma distributions were also used in comparison to the histogram of the total radiance and the Statistics Toolbox (ver. 5.3.1) of MATLAB was used to implement these distributions. MATLAB's MLE for the gamma and Weibull use a numerical minimization technique based on a Neldor-Meade simplex search algorithm (MATLAB Statistics Toolbox Manual, 1999). The gamma probability distribution function (pdf) is

$$\frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta},$$ (4.3)

where α and β are the parameters of the distribution. Gamma pdfs are skewed left with long right tails and are non-negative. The Weibull probability distribution function is

44

$$\frac{\alpha}{\beta^{\alpha}} x^{\alpha-1} e^{-(x/\beta)^{\alpha}} , \qquad\qquad (4.4)$$

where α and β are the parameters of the distribution.

The Weibull distribution was explored in this research since it proved useful in RADAR applications (Stein, 1997). In contrast to the other distributions, the Weibull distribution can be skewed in either direction. However, this distribution did not provide a good fit to most of the data in this study, and was difficult to implement. In some of the following figures, attempts were made to model this distribution, but overall the results were not promising. In some cases, implementation of the MLE Weibull failed to converge to accurate results, and was therefore omitted.

The χ^2 (chi-squared) goodness-of-fit test compares the number of observed events in a particular range of values to the number of expected observations as predicted by a particular probability distribution function. The first step in the chi-squared test is to divide the predictive pdf into approximately equal probability data cells. The next step is to compute the number of predicted and actual observations in each range. The last step is to compute the actual value of the chi-squared statistic and it is given by:

$$\chi^2 = \sum_{\text{all cells}} \frac{(\text{actual observations} - \text{expected observations})^2}{\text{expected observations}} \qquad (4.5)$$

When the number of observations is large, the statistic χ^2 has approximately a chi-squared distribution with $k-1$ degrees of freedom, where k is the number of data ranges considered, when the null hypothesis (H_o) is true (Devore, 1995). When parameters are estimated, the number of degrees of freedom (v) must be reduced to:

$$v = k - p - 1, \tag{4.6}$$

where p is the number of parameters estimated (2 for all distributions considered in this thesis). The number that is produced from equation 4.5 is called the chi-squared value, which denotes the number on the measurement axis such that α (significance level) of the area under the distribution curve with v degrees of freedom, lies to the right of the chi-squared value. That is, the chi-squared value represents the values for which $100*\alpha\%$ of random sequences generated by the distribution would have a *worse* fit than the actual data. Basically, the lower the value of χ^2 in Equation 4.5, the better a certain statistical distribution fits the data.

B.   STATISTICAL MODELING

1.   Total Radiance Analysis for October and June Overflights

A first step in the analysis of the data is to examine the distribution of the total radiance in the images. It is expected that a significant portion of the variability in spectral imagery is due to illumination variations resulting from spatially changing atmospheric effects, shading, shape, and other physical properties of the image. Removal of these effects is an important first step in HSI data processing (Kruse, 1988). The histogram for the total radiance within the C-130 ROI and the corresponding fits for the pdfs considered here are presented in Figure 4.1.
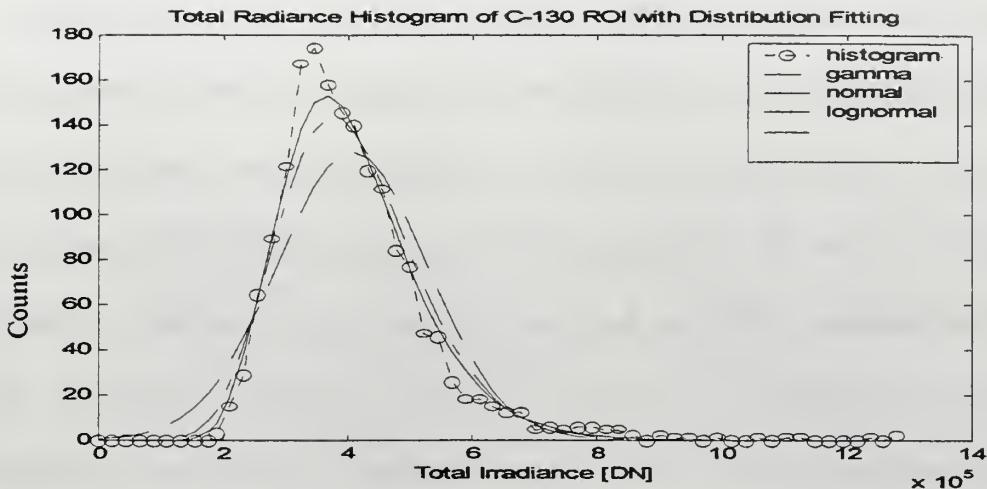
Figure 4.1. Total Radiance Histogram of C-130 ROI with Distribution Fitting.

As was seen in Chapter III, the distribution of the total pixel radiance in the C-130 ROI is highly asymmetric, a normal distribution does not approximate the function well, and the MLE for the Weibull did not converge. The gamma and lognormal distributions capture more of the functional shape of the histogram, but even the lognormal is not quite skewed enough to fit the histogram.

A χ^2 goodness-of-fit test was then performed on the total radiance data from the C-130 ROI. The χ^2 test was described earlier in this chapter and was used here for a visual representation of how well each statistical distribution fits the C-130 data and to determine how many degrees of freedom to use for the statistical analysis later in this chapter. An important part of using the chi-squared test is to choose the right number of degrees of freedom. Too many degrees of freedom may allow poorly fitting distributions to appear acceptable, while too few may have the opposite effect. The objective of the next figure is to demonstrate that the performance of the different distributions is stable across the number of degrees of freedom. In Figure 4.2, curves of equal probability are

47

shown for the chi-squared distribution. These curves give the chi-squared for which $100*\alpha\%$ of random sequences generated by the distribution would have a worse fit than the actual data. As can be seen in the figure, the lognormal is still not ideal, as $\alpha \approx 0.05$. It should be noted that the chi-squared test heavily penalizes spurious observations at the tails of the distribution. While the lognormal distribution has a significance of only 5%, it is clearly a much better fit than the other distributions considered for this ROI. The computed statistical significance of the normal distribution is exactly zero (to floating point precision). Alpha is approximately constant, so 5 degrees of freedom was chosen for the rest of this study.
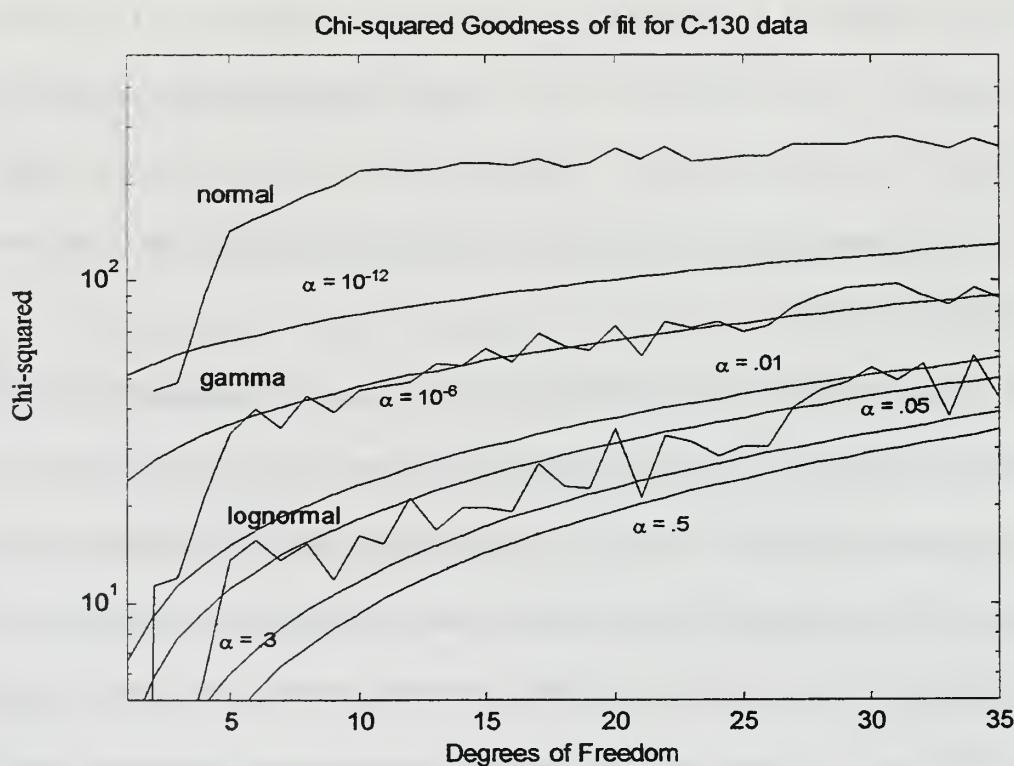

Figure 4.2. Chi-squared values for the distributions of the C-130 ROI.

Table 4.1 shows the parameter values of the best distribution fit for the C-130 ROI. The statistical significance (α) is fairly low for the lognormal, but still better than both the gamma and normal distributions.

Distribution	μ	σ^2	χ^2	α
Lognormal	13.49	.1095	11.03	.05
Gamma	-	-	26.29	0
Normal	7.7×10^5	8.67×10^{10}	113.41	0

Table 4.1. Parameters for distribution fits for C-130 ROI.
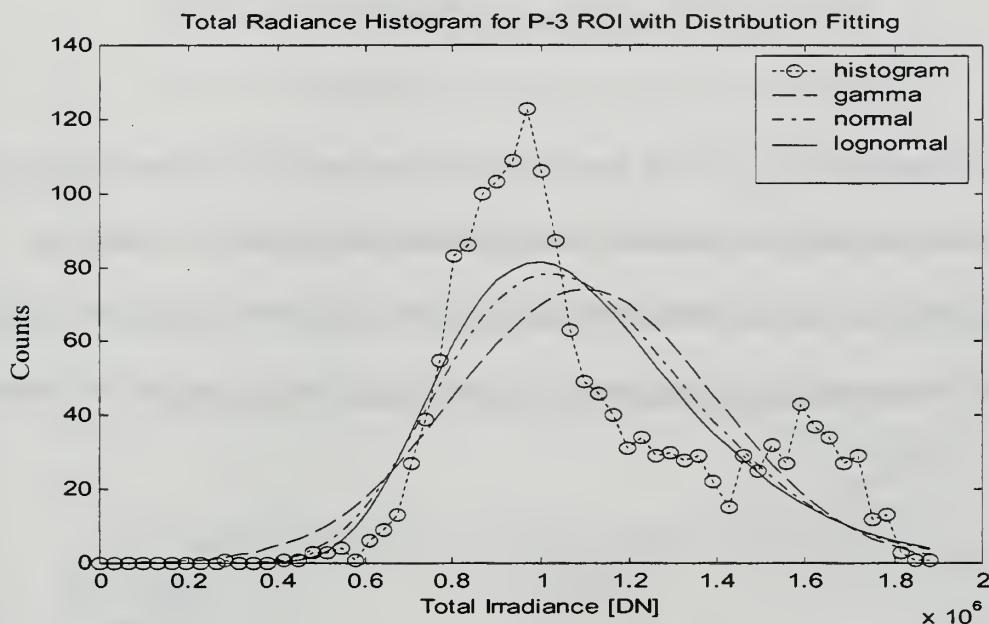
Figure 4.3. Total radiance histogram for P-3 ROI with distribution fitting.

Figure 4.3 shows the P-3 ROI and the different distribution models. As discussed in Chapter III, the histogram of the P-3 ROI is bi-modal, which makes it nearly impossible to model adequately. As is evident in Figure 4.3, none of the distributions provide a proper fit for this ROI, as should be expected due to their uni-modal characteristics. For such problematic data, more advanced methods such as non-

parametric statistics may have to be employed (Devore, 1988). Table 4.2 shows the parameters for the lognormal, which was the best fit, although statistically insignificant, and the gamma and normal distributions.

Distribution	μ	σ^2	χ^2	α
Lognormal	13.45	.0200	221.4	0
Gamma	-	-	245.46	0
Normal	7.0×10^5	1.1×10^{10}	370.70	0

Table 4.2. Parameters for distribution fits for P-3 ROI.

In Chapter III, the P-3 ROI was broken into two parts, the fuselage and the wings. It was shown that the main contributor to the bi-modal nature of the P-3 ROI total radiance histogram was anomalies in the fuselage data. Figure 4.4 displays the total radiance histogram of the aircraft wings from the P-3 ROI with respective distribution fits.
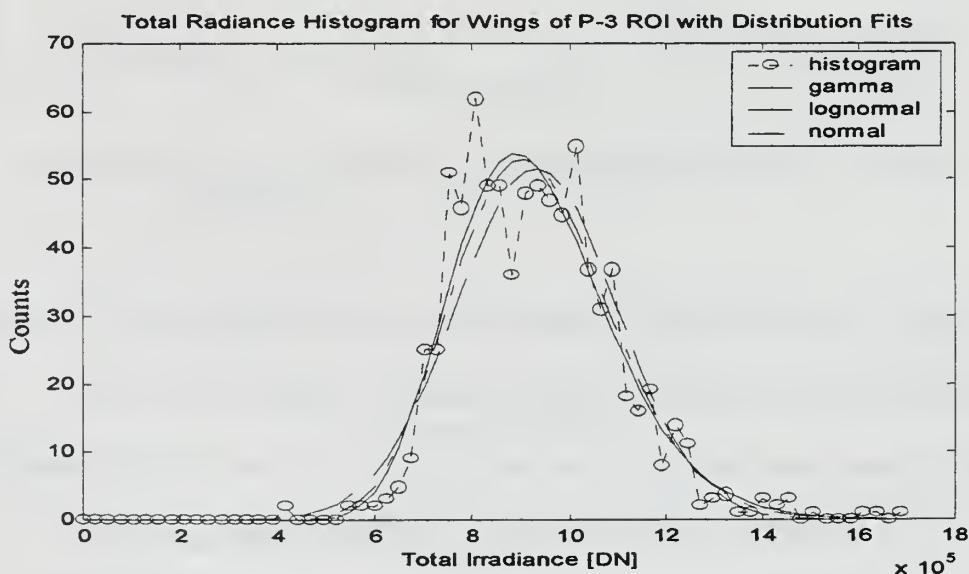

Figure 4.4. Total Radiance Histogram for Wings in P-3 ROI with Dist. Fits.

50

The model statistical distributions fit much better with the wing data, as they should. The lognormal distribution fit the P-3 wing histogram the best. Since the bi-modal characteristics of the P3 ROI prohibited accurate modeling with statistical distributions, further analysis of this ROI will be concentrated on the wing data. Table 4.3 lists the parameters for the distributions.

Distribution	μ	σ^2	χ^2	α
Lognormal	13.46	.0219	106.04	0
Gamma	-	-	114.80	0
Normal	7.1×10^5	1.14×10^{10}	159.32	0

Table 4.3. Parameters for distribution fits for P-3 Wing ROI.

Figure 4.5 shows the Fence line ROI data distribution with the respective statistical models. The histogram is slightly asymmetric with a left tail. All three models
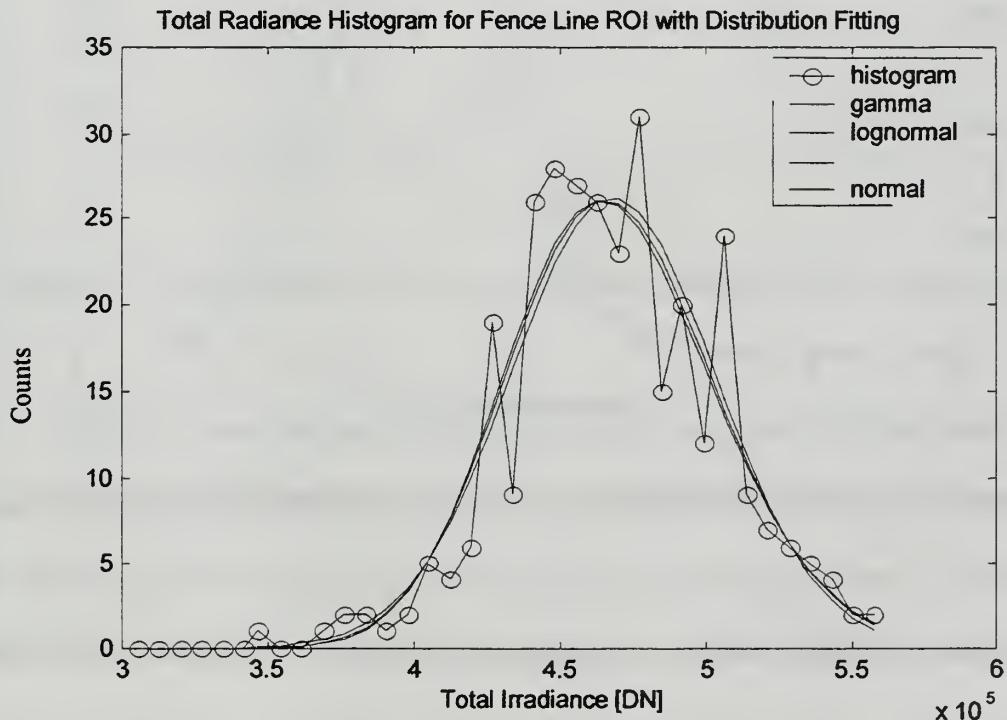
Figure 4.5. Total Radiance Histogram for Fence Line ROI with Distribution Fitting

51

are nearly identical in shape in this figure, and capture some of the histogram's shape. Of

the three models, the normal was the best fit for this histogram. Table 4.4 shows the

parameters for the distribution fits.

Distribution	μ	σ^2	χ^2	α
Normal	6.4×10^5	2.3×10^{10}	7.67	.10
Lognormal	13.34	.0581	23.25	.0002
Gamma	-	-	16.81	.0054

Table 4.4. Parameters for distribution fits for Fence Line ROI.
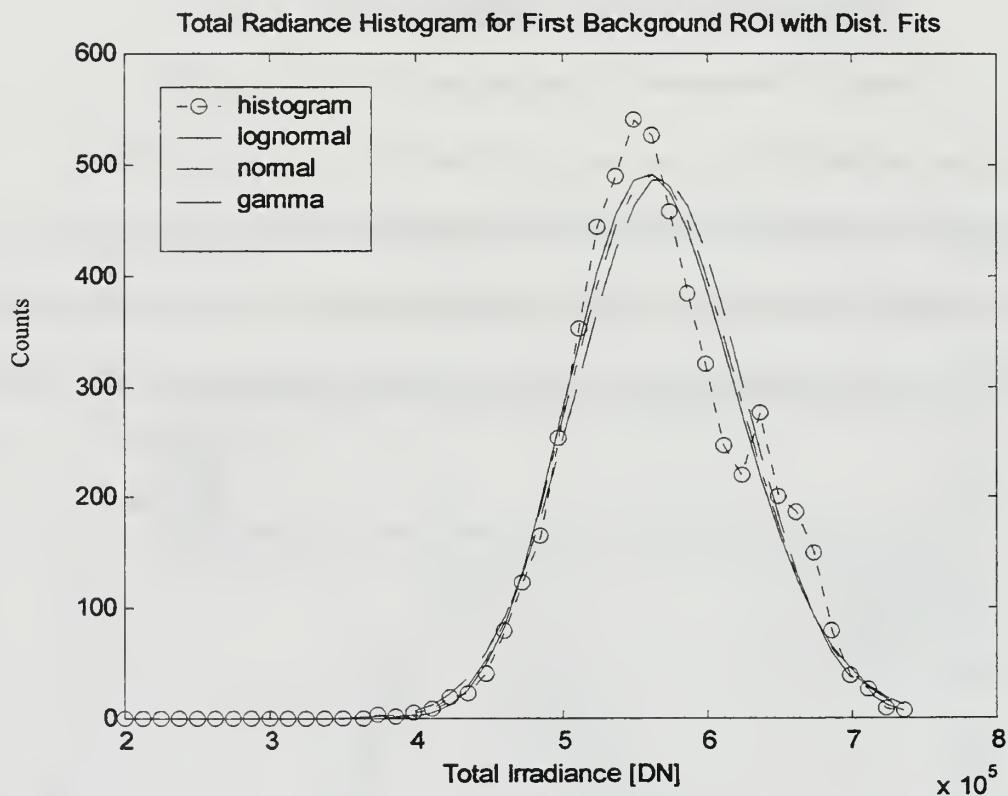
Figure 4.6. Total Radiance Histogram for First Background ROI with Dist. Fits

Figure 4.6 displays the first background ROI and the fitted statistical distributions.

The histogram is slightly skewed to the right, and all three statistical models appeared to fit somewhat well. The parameter values are shown in Table 4.5. The lognormal had the lowest chi-squared value of the three distributions, but was statistically insignificant.

Distribution	μ	σ^2	χ^2	α
Lognormal	13.43	.0260	138.3	0
Gamma	-	-	866.8	0
Normal	6.9×10^5	9.2×10^{10}	569.5	0

Table 4.5. Parameters for distribution fits of First Background ROI.
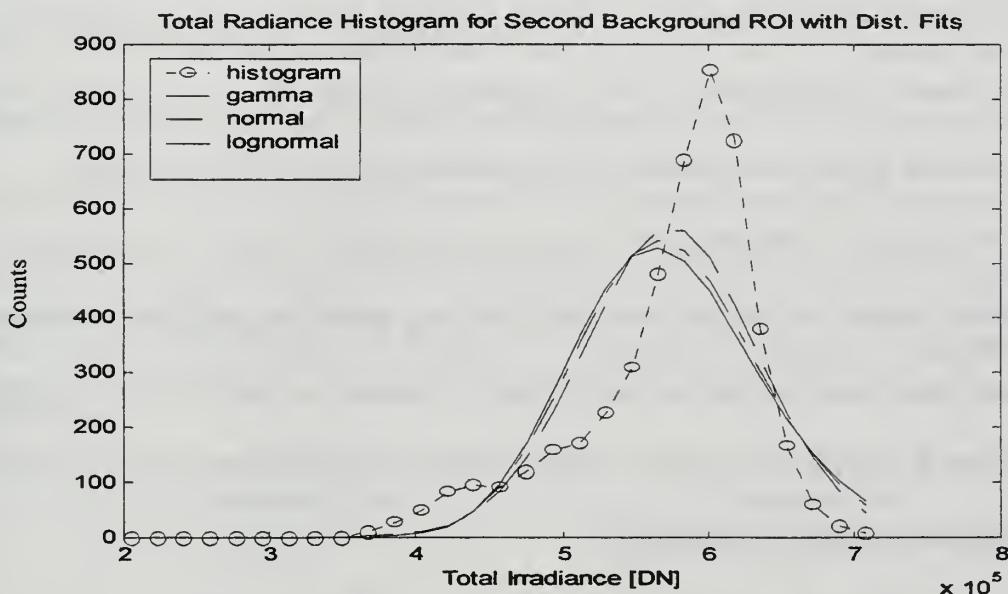
Figure 4.7. Total Radiance Histogram for Second Background ROI with Dist. Fits.

Figure 4.7 shows the second background ROI histogram with fitted distributions. None of the distributions fit this histogram due to its right skewed nature and large peak. The reason for this is that the above histogram has a left tail and you need a distribution with a left tail to model it. The gamma and lognormal plots have right tails while the normal is symmetric. The Weibull was included because of its ability to model such shapes, but as discussed earlier, there were implementation difficulties. Table 4.6 shows

53

the parametric data from the lognormal, which, while it was the best of the three
distributions, was still statistically insignificant.

Distribution	μ	σ^2	χ^2	α
Lognormal	13.44	.1164	2262	0
Gamma	-	-	2992	0
Normal	7.3×10^5	7.5×10^{10}	3746	0

Table 4.6. Parameters for Best Distribution fit.

The next step in the analysis of the total radiance data from the various ROIs is
the use of quantile-quantile (qq) plots. Plotting two rank-ordered vectors corresponding
to random observations against each other composes a qq plot. The (x, y) position of the
first point on the qq plot is determined by the minimum values of the two random
vectors, the last point is determined by their maximum values, and so on in between.
Two different random vectors that come from the same underlying statistical distribution
result in an approximately linear qq plot. Figure 4.8 presents qq plots for the C-130 ROI.
The ROI data is plotted as X-quantiles and the random variables drawn from the various
distributions are plotted as Y-quantiles.
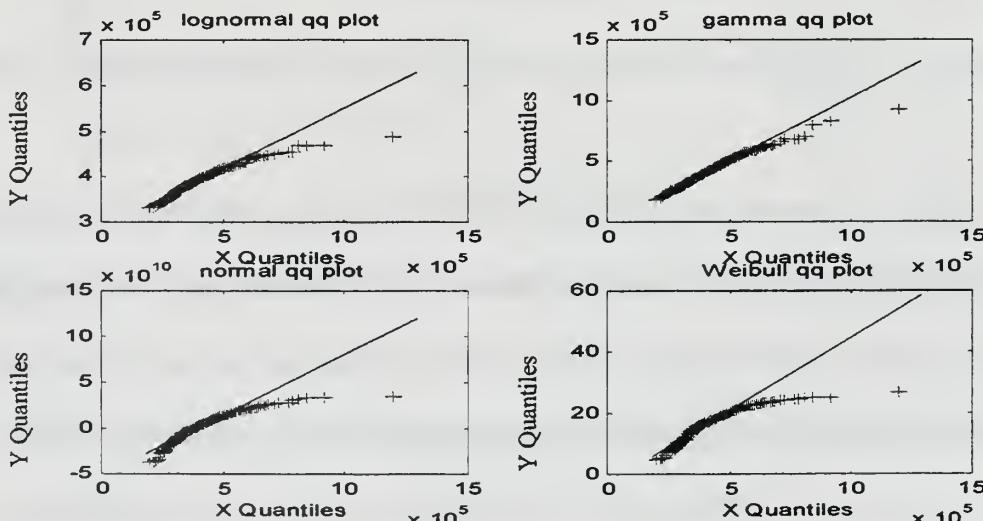
Figure 4.8. QQ Plots for C-130 ROI.

54

The qq plot helps to determine where the problem lies with a particular distribution. For example, in Figure 4.8 none of the distributions predict enough high-end observations since the plotted points fall below the linear dashed line at the high end, and only the gamma, and to a lesser extent the lognormal have appropriate lower tails. The normal and Weibull plots from Figure 4.8 have too many low-end observations since their plots are below the linear dashed line. Basically, for the lower end of the qq plot, if observations are plotted below the line (Weibull plot in Figure 4.9), then there are too many low-end observations and if observations are plotted above (Lognormal plot in Figure 4.10), then there are not enough low-end observations. For the upper end of the qq plot, if observations are plotted below the linear dashed line (Weibull plot in Figure 4.9), then that means that there is not enough high-end observations and if observations are plotted above (Gamma plot in Figure 4.10), then there are too many high-end observations.
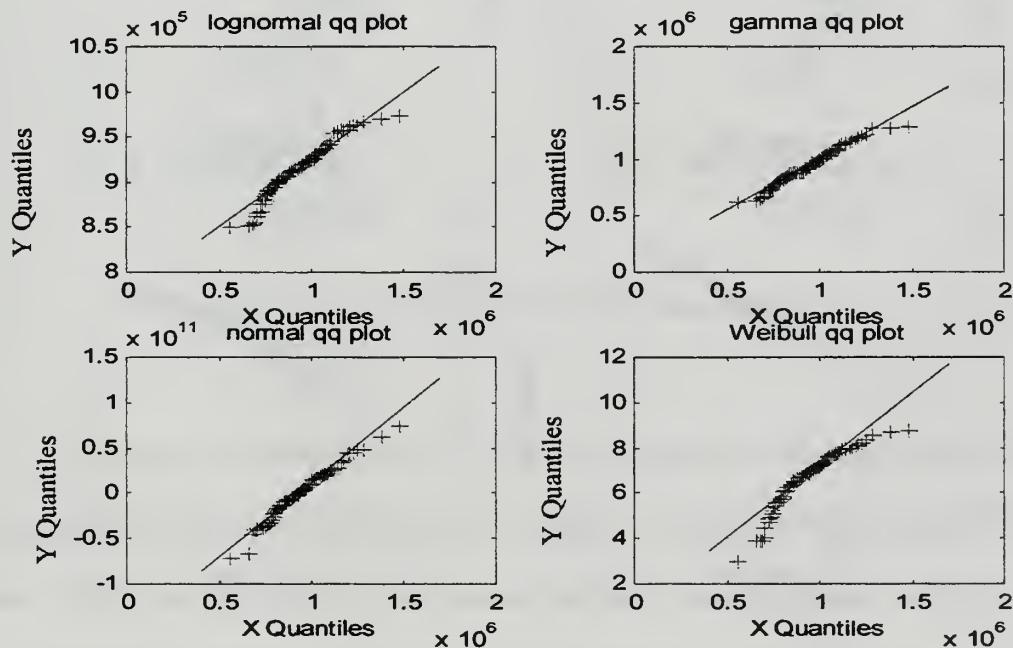
Figure 4.9. QQ plots for Wings of P-3 ROI.

55

The qq plots presented in Figure 4.9 are for wings of the P-3 ROI. The gamma qq plot seems to be the most linear in nature, although the normal qq plot was fairly linear as well. This means that the data is likely to be described by either a gamma or a normal distribution.

Figure 4.10 shows the qq plots for the fence line ROI. The lognormal and normal plots did not predict enough low-end observations, while the Weibull predicted too many low-end observations. The gamma plot predicted too many high-end observations. The normal distribution appears to model the data best, however, some of its lower tail predictions need to be more linear to adequate model this data.
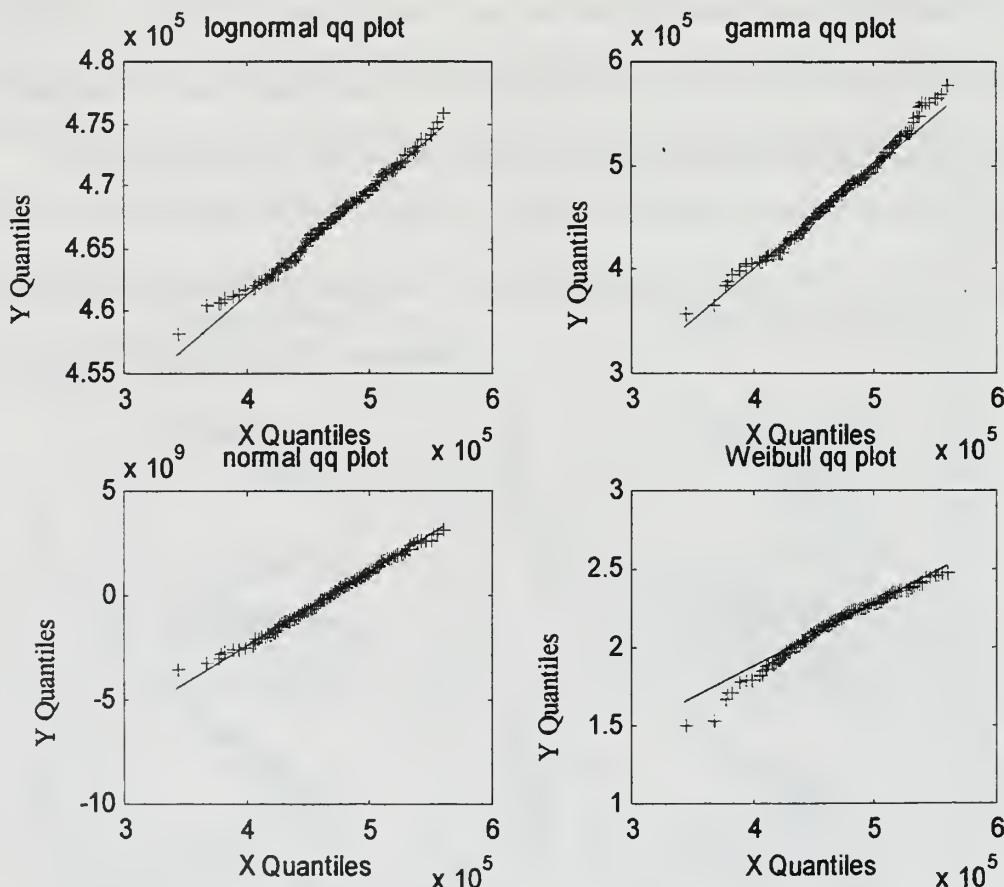
Figure 4.10. QQ plots for Fence Line ROI.

Figure 4.11 displays the qq plots for the first background ROI. The lognormal appears to model this background data the best. The lower and upper tails of the lognormal plot need to be slightly more linear for this distribution to be an accurate model. For the most part, all four distributions were fairly linear with only slight deviations.
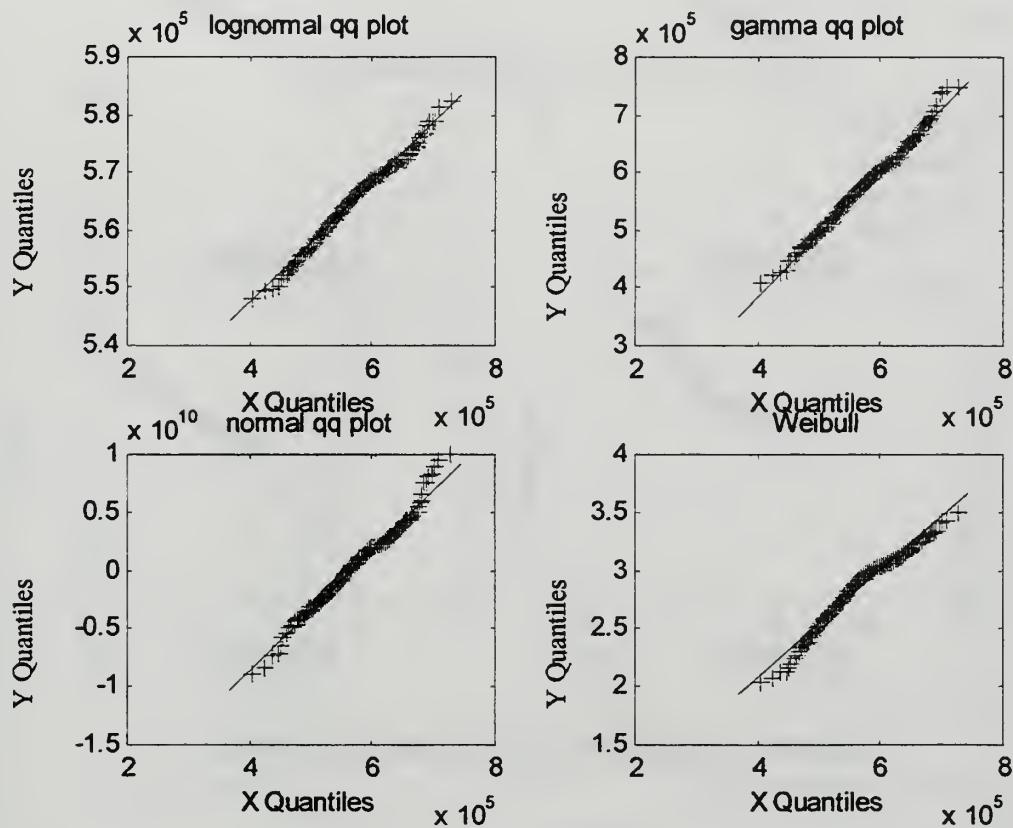
Figure 4.11. QQ plots for First Background ROI.

The last qq plot displayed here is the plot of the second background ROI, which is predominately composed of sand and dirt. Figure 4.11 shows the four qq plots for this data. None of the four distributions does an adequate job of modeling this ROI data. The lower tails of these distributions do not predict the lower-end observations at all, probably

due to the highly skewed nature of this background data. The appearance of the qq plots for the second background ROI is significantly different than the qq plots of the first background ROI.
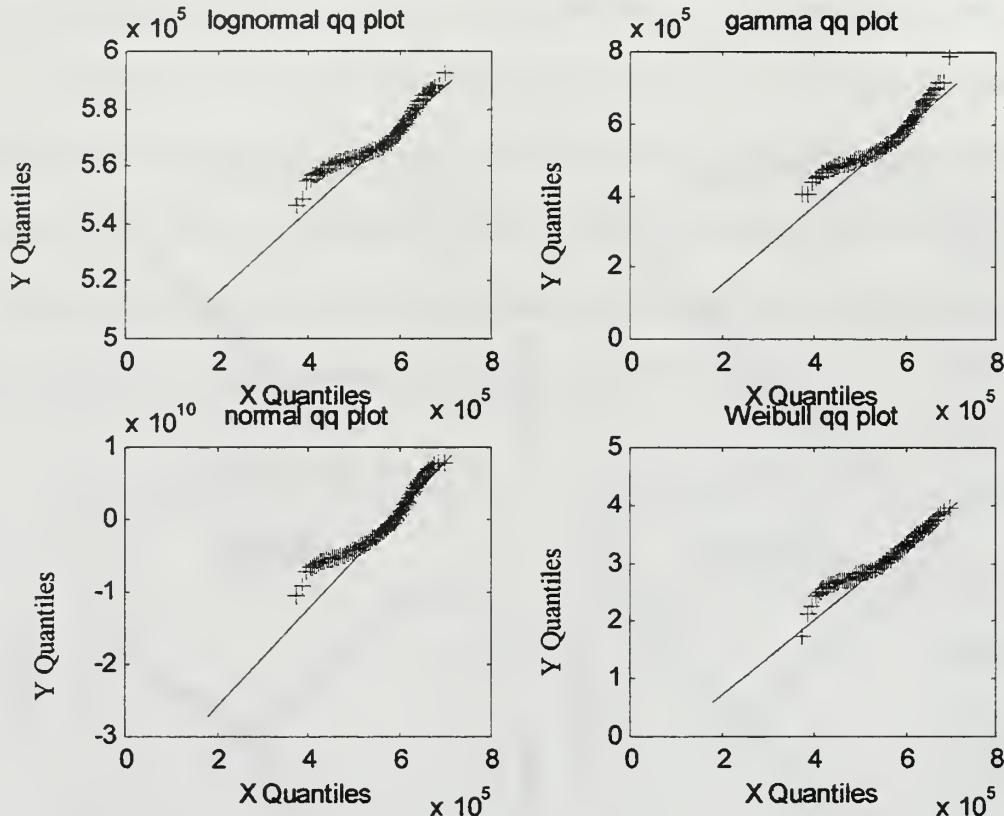
Figure 4.11. QQ plots for Second Background ROI

As a comparison to the October over-flight data, a detailed statistical analysis was also done on the regions of interest from the June over-flight data. The first total radiance histogram with distribution fits is of the C-130 ROI data and is shown in Figure 4.12. The results of the statistical distribution fits are the same as the fits done on the C-130 data from the October over-flight. The lognormal distribution fit the total radiance

histogram the best of the statistical distributions tested here and the parameters are shown

in Table 4.7.

Distribution	μ	σ^2	χ^2	α
Lognormal	13.14	.0843	8.76	.31
Gamma	-	-	17.47	.006
Normal	5.3×10^5	2.5×10^{10}	71.66	0

Table 4.7. Parameters for distribution fits of June C-130 ROI.
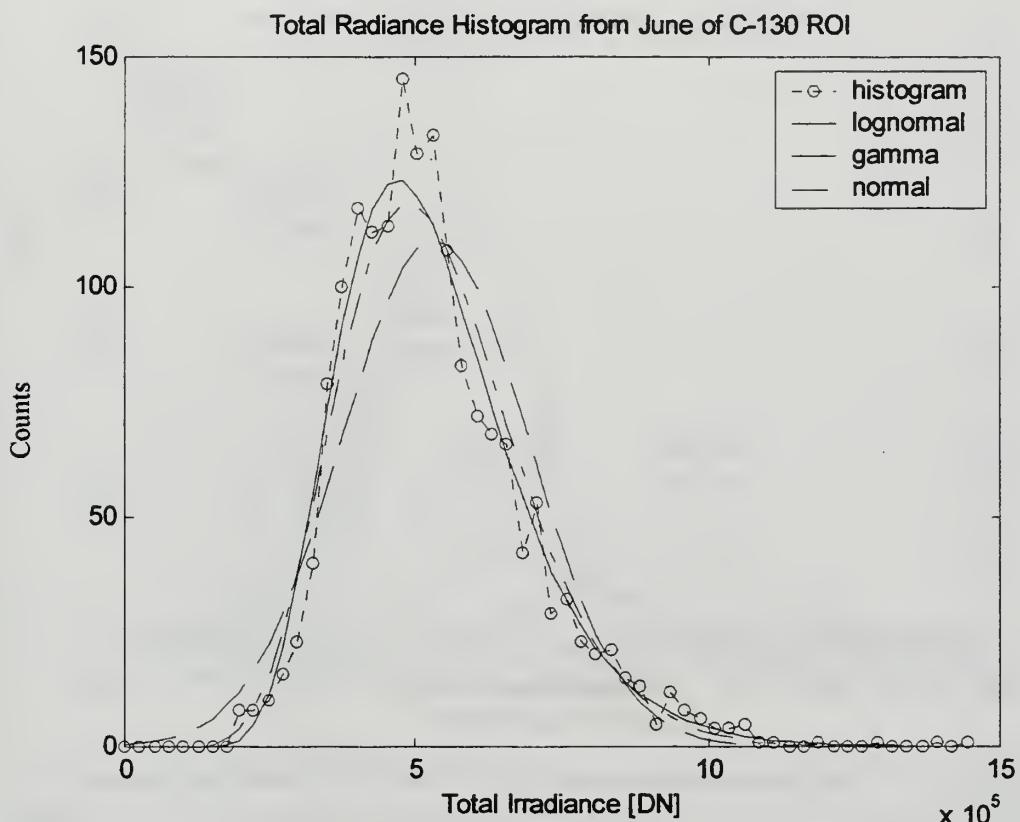
Figure 4.12. Total radiance histogram of June C-130 data with Dist. Fits.

Figure 4.13 displays the total radiance histogram of the P-3 Wing ROI from the

June over-flight. Again, the results are very similar to the P-3 Wing data from the

October over-flight which was shown in Figure 4.4. The lognormal distribution fit the

June P-3 wing data best of the statistical distributions, and the results are shown below in

Table 4.8.

Distribution	μ	σ^2	χ^2	α
Lognormal	13.97	.0254	5.31	.65
Gamma	-	-	5.47	.60
Normal	1.2×10^6	3.45×10^{10}	11.74	.052

Table 4.8. Parameters for distribution fits of June P-3 Wing ROI.
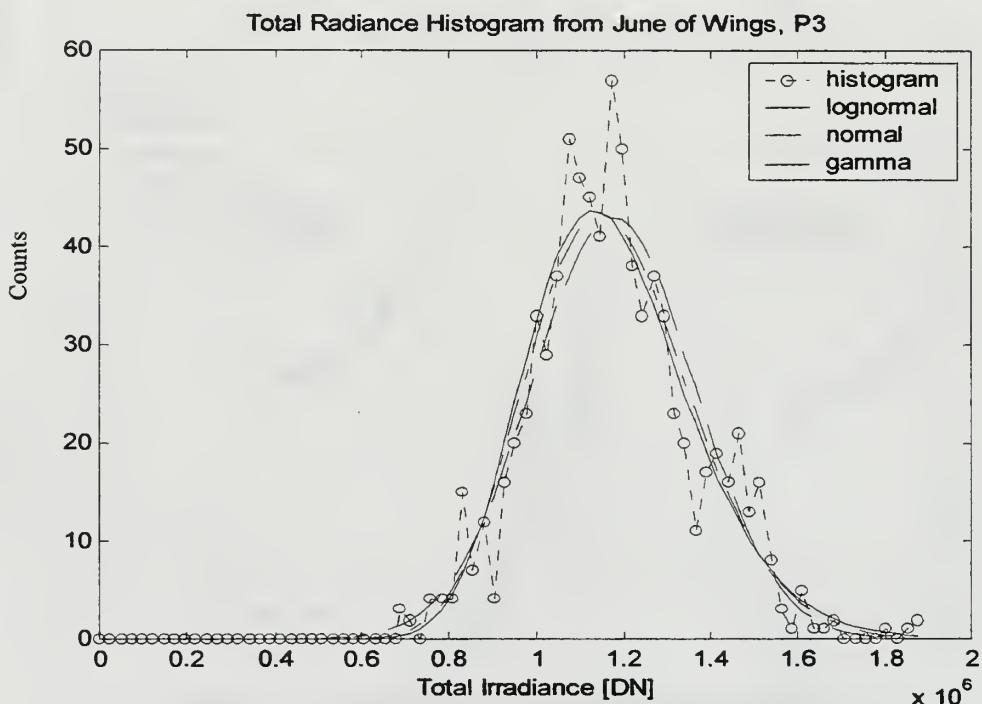
Figure 4.13. Total radiance histogram of P-3 Wing ROI with distribution fits.

2.   Single Band Statistics for October and June Overflights

The investigation of the distribution of total radiance provides information about

overall intensity distributions within a ROI. The next important step is to analyze the

statistics associated with individual bands, specifically the unnormalized single band

radiance data for each of the regions of interest. It is expected that some of the effects of

60

albedo that were removed through the normalization procedure will be reintroduced into the spectral data and produce more variance. Figure 4.14 displays the unnormalized single band histograms from the C-130 ROI.

Unnormalized Single Band Histograms of C-130 ROI with Dist. Fits
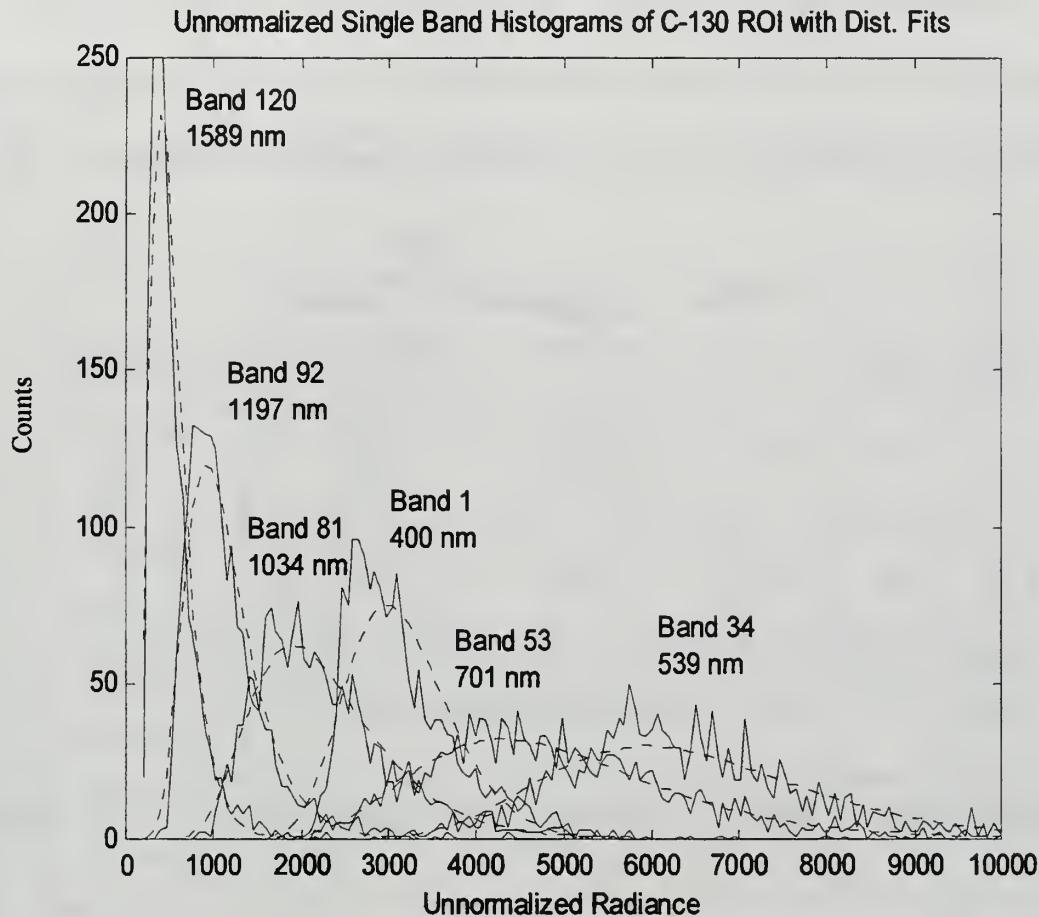
Figure 4.14. Unnormalized single band histograms of C-130 ROI with Dist. Fits.

From Figure 4.14, it seems that the statistical distributions are able to adequately model the histograms. The lognormal distribution fit all six histograms best. After applying the χ^2 test with the different statistical distributions, Band 53 showed a chi-squared value of 1.86 with a significance level of .8387 for the lognormal, which was the best of the six

bands. Band 81 also had a low value of 7.04 with a significance level of .2178 for the

lognormal. The best fitting distributions and corresponding parameters are given in

Table 4.9. The data indicates that no one distribution is adequate for all of the single

band, unnormalized radiance histograms. The χ^2 goodness-of-fit test was used, with 5

degrees of freedom, for the various histograms plotted in Figure 4.14. The parameters in

the last column correspond to the predicted histograms also plotted in Figure 4.14.

		Normal		Lognormal		Gamma		
Band	λ(nm)	χ^2	α	χ^2	α	χ^2	α	Best Params
1	400	325.9	0	140.0	0	184.5	0	$\mu_{ln} = 8.04$, $\sigma_{ln}^2 = 0.037$
34	539	171.8	0	34.59	0	62.96	0	$\mu_{ln} = 8.75$, $\sigma_{ln}^2 = 0.058$
53	701	102.5	0	1.86	.839	12.87	.030	$\mu_{ln} = 8.46$, $\sigma_{ln}^2 = 0.091$
81	1034	225.8	0	7.04	.218	44.39	0	$\mu_{ln} = 7.68$, $\sigma_{ln}^2 = 0.120$
92	1197	343.7	0	23.36	.0003	99.51	0	$\mu_{ln} = 6.96$, $\sigma_{ln}^2 = 0.138$
120	1589	419.1	0	53.01	0	133.6	0	$\mu_{ln} = 6.18$, $\sigma_{ln}^2 = 0.183$

Table 4.9. Goodness of fit for the various histograms presented in Figure 4.14 for the C-130 ROI. The Weibull distribution was not always able to converge and was therefore omitted.
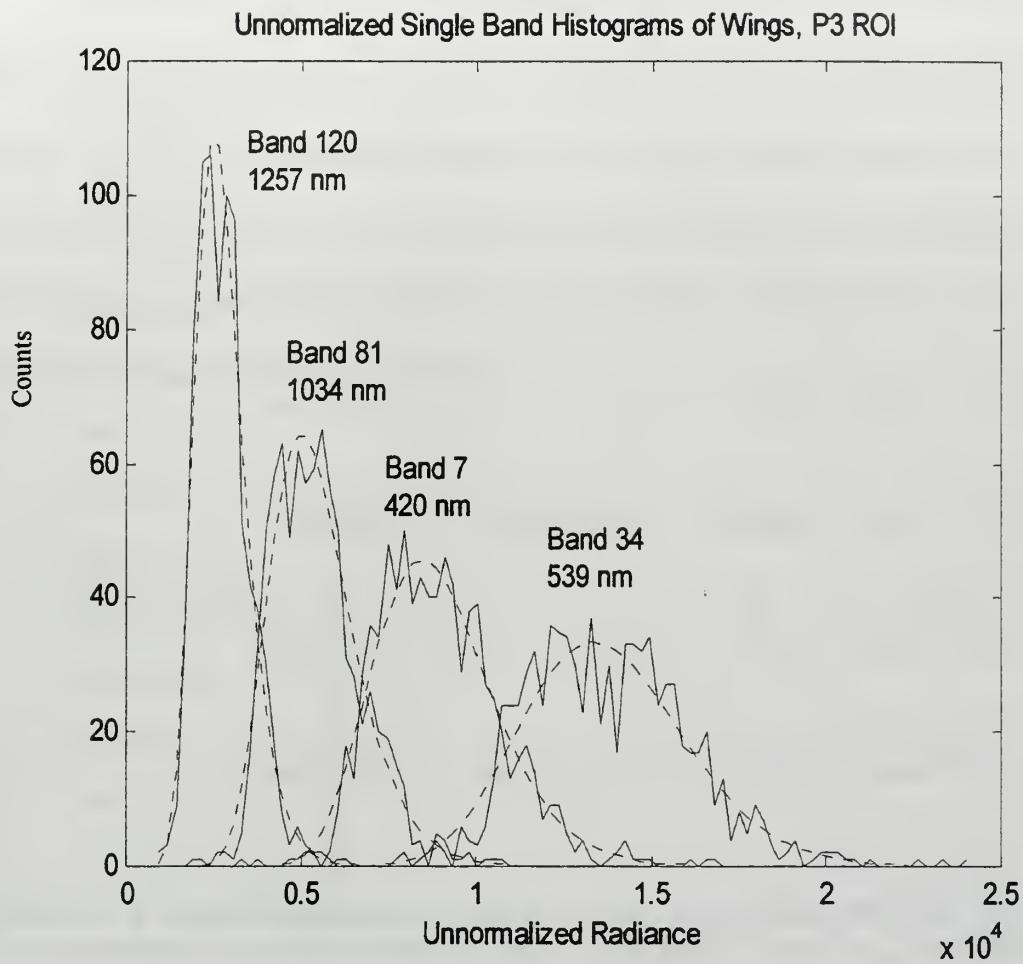
Figure 4.15. Unnormalized single band histograms of P-3 Wing ROI with distribution fits.

Figure 4.15 shows the unnormalized single band histograms for the P-3 Wing data. For a better visual depiction of the histograms, the histograms of Bands 59 and 84 were not displayed in the above figure. The lognormal distribution fit best for all six bands. The parameter values for the distribution fits of this data are listed in Table 4.10.

		Normal		Lognormal		Gamma		
Band	λ(nm)	χ^2	α	χ^2	α	χ^2	α	Best Params
7	420	25.29	0	22.72	.0004	22.85	.0004	$\mu_{ln} = 9.09,$ $\sigma_{ln}^2 = 0.026$
34	539	16.16	.005	12.14	.033	13.05	.027	$\mu_{ln} = 9.51,$ $\sigma_{ln}^2 = 0.029$
59	742	28.33	0	2.99	.70	7.99	.236	$\mu_{ln} = 9.08,$ $\sigma_{ln}^2 = 0.038$
81	1034	42.83	0	12.18	.032	18.65	.003	$\mu_{ln} = 8.40,$ $\sigma_{ln}^2 = 0.057$
84	1078	44.97	0	8.33	.139	11.30	.042	$\mu_{ln} = 7.94,$ $\sigma_{ln}^2 = 0.068$
96	1257	35.07	0	13.47	.019	16.65	.009	$\mu_{ln} = 7.89,$ $\sigma_{ln}^2 = 0.078$

Table 4.10. Goodness of fit for the various histograms presented in Figure 4.15 for the P-3 Wing ROI. The Weibull distribution was not always able to converge and was therefore omitted.
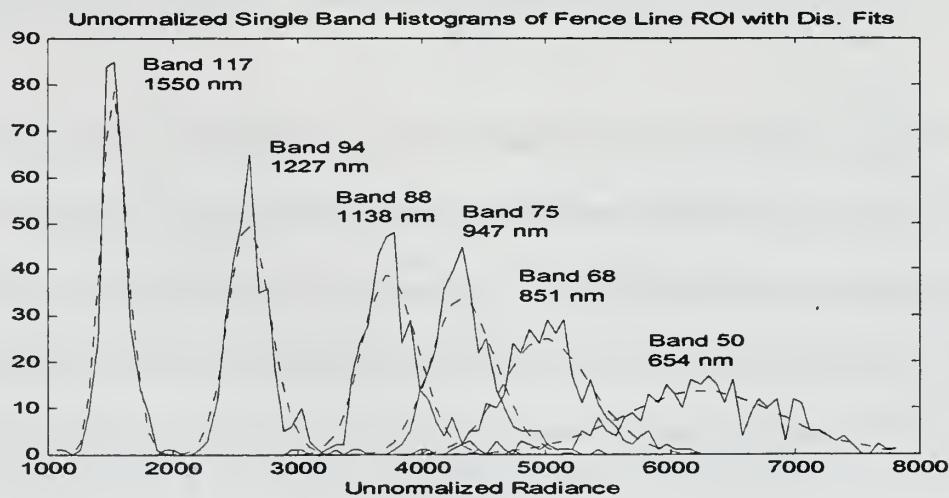

Figure 4.16. Unnormalized single band histograms of Fence Line ROI with dist. fits.

Figure 4.16 displays unnormalized single band histograms from the Fence Line ROI. The histograms appear symmetric and the normal distribution fits three of the band histograms while the lognormal fits the other three. The goodness-of-fit for these distributions is very good, with an average significance level of .371 for the normal distribution fit and an average significance level of .122 for the lognormal fit. The parameter values are shown in Table 4.11.

		Normal		Lognormal		Gamma		
Band	λ(nm)	χ^2	α	χ^2	α	χ^2	α	Best Params
50	654	7.27	.201	10.13	.072	10.95	.052	$\mu_n = 4374$, $\sigma_n^2 = 1.4 \times 10^5$
68	851	6.07	.299	7.00	.220	6.41	.268	$\mu_n = 5629$, $\sigma_n^2 = 3.3 \times 10^5$
75	947	3.57	.6126	5.26	.385	5.05	.410	$\mu_n = 6250$, $\sigma_n^2 = 4 \times 10^5$
88	1138	21.51	.0006	17.26	.004	18.29	.0026	$\mu_{ln} = 8.38$, $\sigma_{ln}^2 = 0.035$
94	1227	10.57	.061	7.72	.1722	9.63	.086	$\mu_{ln} = 7.87$, $\sigma_{ln}^2 = 0.0045$
117	1550	14.46	.013	7.43	.1906	12.62	.027	$\mu_{ln} = 7.89$, $\sigma_{ln}^2 = 0.0064$

Table 4.11. Goodness of fit for the various histograms presented in Figure 4.16 for the Fence Line ROI. The Weibull distribution was not always able to converge and was therefore omitted.
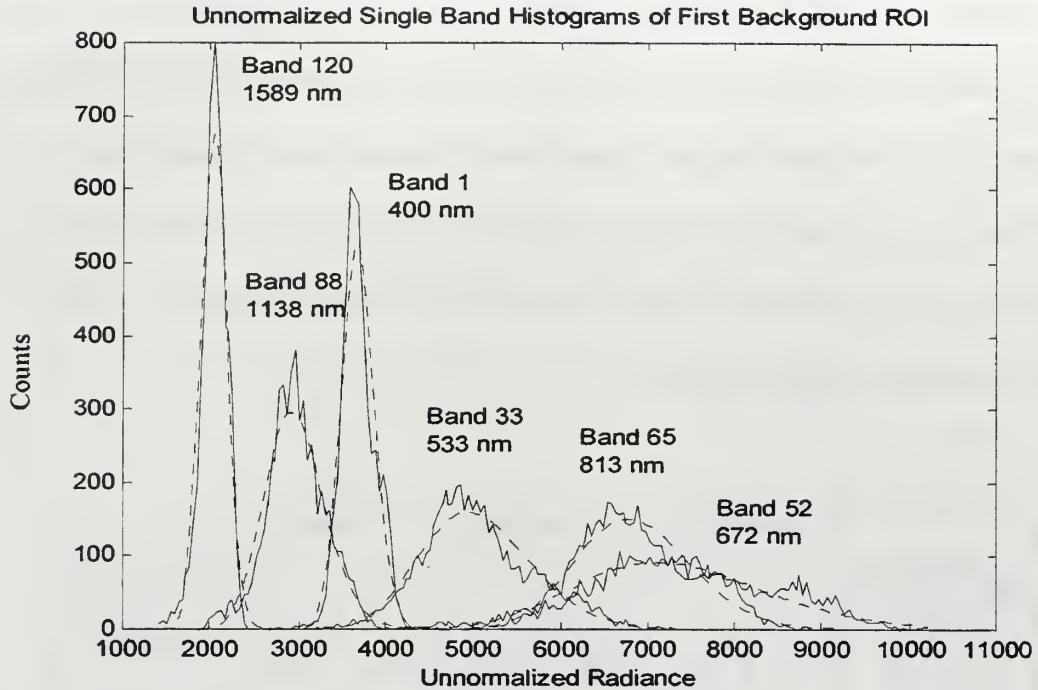
Figure 4.17. Unnormalized single band histograms of First Background ROI with best fitting distributions.

Figure 4.17 displays the unnormalized single band histograms for the First Background ROI and the corresponding best-fit distributions. The histograms in Figure 4.17 appear to be much different from the normalized single band histograms of this ROI, that will be shown in the next section. The histograms of Bands 33, 52, and 65 are more spread out along the radiance axis. The lognormal distribution was the best fit for all of the band histograms except for Band 120, which was fit best by the normal. The χ^2 test revealed values in excess of 65 and significance levels of zero which indicates that the distribution fits for this ROI are not significant. Table 4.12 shows the parameter values of the different distributions and results of the goodness-of-fit tests.

		Normal		Lognormal		Gamma		
Band	λ(nm)	χ^2	α	χ^2	α	χ^2	α	Best Params
1	400	100.8	0	65.31	0	66.89	0	$\mu_{ln} = 7.99,$ $\sigma_{ln}^2 = 0.013$
33	533	232.7	0	106.5	0	143.6	0	$\mu_{ln} = 8.52,$ $\sigma_{ln}^2 = 0.016$
52	672	92.50	0	66.82	0	68.33	0	$\mu_{ln} = 8.89,$ $\sigma_{ln}^2 = 0.023$
65	813	231.6	0	122.7	0	158.3	0	$\mu_{ln} = 8.83,$ $\sigma_{ln}^2 = 0.009$
88	1138	255.1	0	193.1	0	212.9	0	$\mu_{ln} = 8.13,$ $\sigma_{ln}^2 = 0.003$
120	1589	139.6	0	154.2	0	156.6	0	$\mu_n = 1906,$ $\sigma_n^2 = 1.5 \times 10^4$

Table 4.12. Goodness of fit for the various histograms presented in Figure 4.17 for the First Background ROI. The Weibull distribution was not always able to converge and was therefore omitted.

The next figure will display the unnormalized radiance, single band data from the Second Background ROI. Figure 4.18 shows the histograms and corresponding best-fit statistical distributions. As was the case with the First Background ROI, the histograms of Bands 33, 52, and 65 are much different in shape from the normalized radiance histograms of this data. The statistical distributions do not appear to fit well, and the χ^2 test confirmed that fact with values exceeding 180 and significance values of zero. The normal distribution fit best for these histograms, although it was statistically insignificant. Table 4.13 displays the parameters of the distribution fits for this ROI.
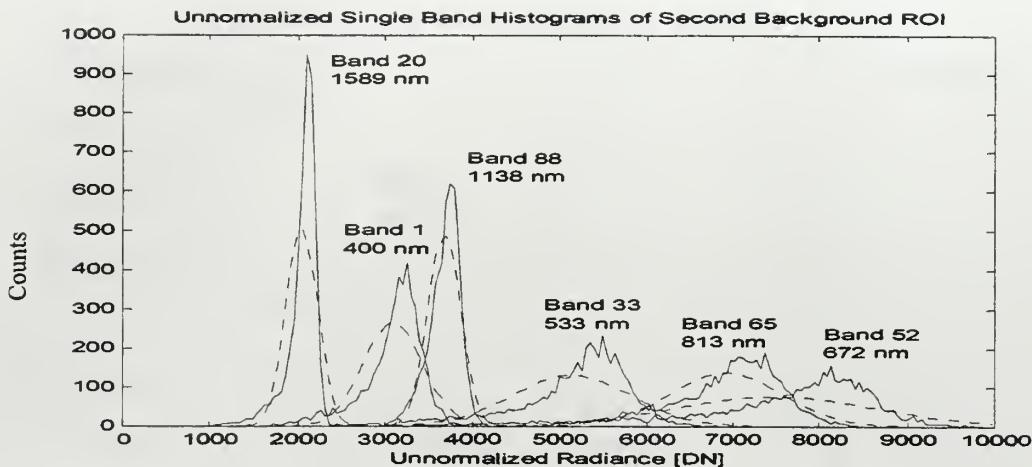
Figure 4.18. Unnormalized single band histograms of Second Background ROI with best fitting distributions.

Band	λ(nm)	Normal		Lognormal		Gamma		
		χ^2	α	χ^2	α	χ^2	α	Best Params
1	400	420.5	0	825.5	0	570.7	0	$\mu_n = 3096$, $\sigma_n^2 = 1.1 \times 10^5$
33	533	589.9	0	1071	0	792.8	0	$\mu_n = 3144$, $\sigma_n^2 = 4.7 \times 10^5$
52	672	653.1	0	1028	0	864.2	0	$\mu_n = 7497$, $\sigma_n^2 = 1.4 \times 10^6$
65	813	259.7	0	400.0	0	351.5	0	$\mu_n = 6919$, $\sigma_n^2 = 4.3 \times 10^5$
88	1138	186.9	0	298.1	0	247.7	0	$\mu_n = 3387$, $\sigma_n^2 = 2.8 \times 10^4$
120	1589	810.7	0	1845	0	1632	0	$\mu_n = 1889$, $\sigma_n^2 = 2.1 \times 10^4$

Table 4.13. Goodness of fit for the various histograms presented in Figure 4.18 for the Second Background ROI. The Weibull distribution was not always able to converge and was therefore omitted.

The following section displays the normalized single band histograms for each of the ROIs from the October over-flight. As was discussed in Chapter III, the simple procedure of normalization is performed to eliminate the effects of albedo in processing spectral data. A representative sample of the histograms, displayed in the previous chapter from the C-130 ROI, is plotted in Figure 4.19. For each of the six bands depicted in Figure 4.19, the four MLE distributions were computed for the functions discussed previously.

Single Band Histograms with Distribution Fitting for C-130 ROI
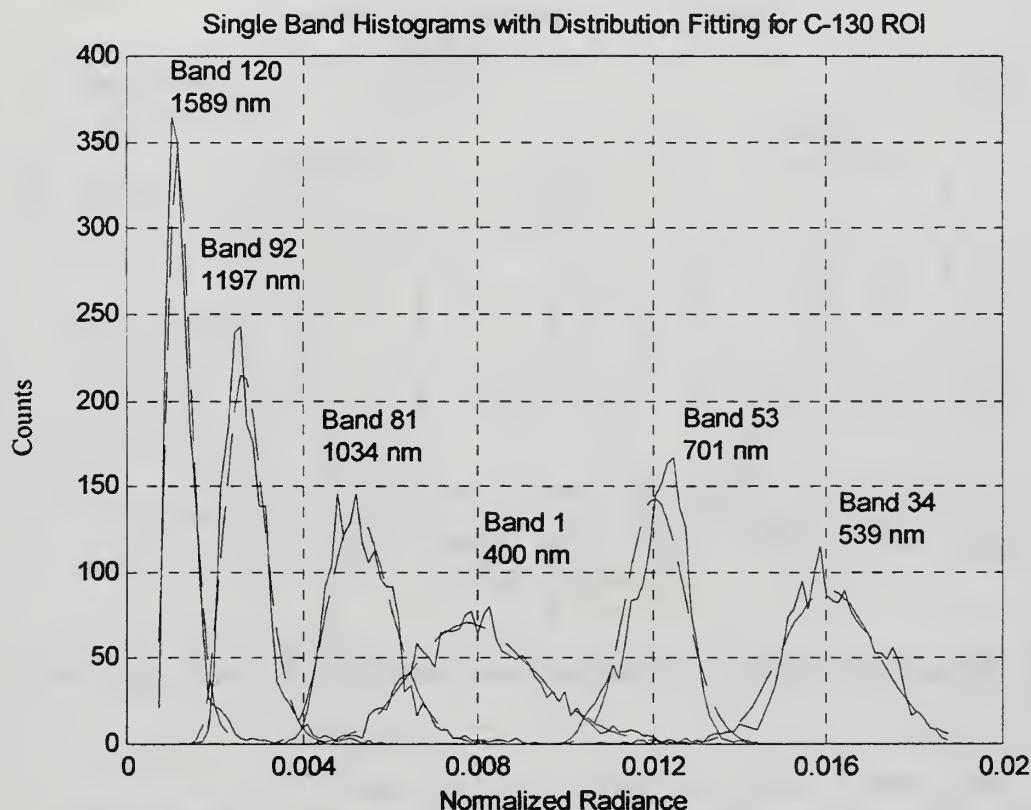
Figure 4.19. Single band histograms with distribution fitting for C-130 ROI.

Along with each of the histograms is plotted the predicted histogram corresponding to the distribution that best fits the data. The gamma distribution fit best

69

for Band 1, the normal fit best for Band 53 and the lognormal distribution fit best for the other four bands. The best fitting distributions and corresponding parameters are given in Table 4.14. The data indicates that no one distribution is adequate for all of the single bands, normalized radiance distributions. The χ^2 goodness-of-fit test was used, with 5 degrees of freedom, for the various histograms plotted above. The parameters in the last column correspond to the predicted histograms also plotted in Figure 4.19.

		Normal		Lognormal		Gamma		
Band	λ(nm)	χ^2	α	χ^2	α	χ^2	α	Best Params
1	400	10.4	0.064	8.55	0.12	3.28	0.66	$\alpha_\Gamma = 32.9,$ $\beta_\Gamma = 2.44 \times 10^{-3}$
34	539	27.1	10^{-4}	24.1	2×10^{-4}	27.0	10^{-4}	$\mu_{ln} = -4.1,$ $\sigma_{ln}^2 = 4.5 \times 10^{-3}$
53	701	45.8	0	59.2	0	54.1	0	$\mu_n = 0.0121,$ $\sigma_n^2 = 4.45 \times 10^{-7}$
81	1034	65.5	0	18.6	.0023	29.6	0	$\mu_{ln} = -5.23,$ $\sigma_{ln}^2 = 0.019$
92	1197	179	0	55.5	0	84.0	0	$\mu_{ln} = -5.92,$ $\sigma_{ln}^2 = 0.029$
120	1589	199	0	51.7	0	83.5	0	$\mu_{ln} = -6.69,$ $\sigma_{ln}^2 = 0.055$

Table 4.14. Goodness of fit for the various histograms presented in Figure 4.19 for the C-130 ROI. MLE Weibull were not always able to converge, and were therefore omitted.

Figure 4.20 shows the single band histograms and corresponding statistical fits for the wings from the P-3 ROI. Because of the bi-modal nature already discussed, only the wing data from the entire P-3 ROI will be displayed here. Figure 4.20 shows the single band histograms and the corresponding fits of the different statistical distributions.
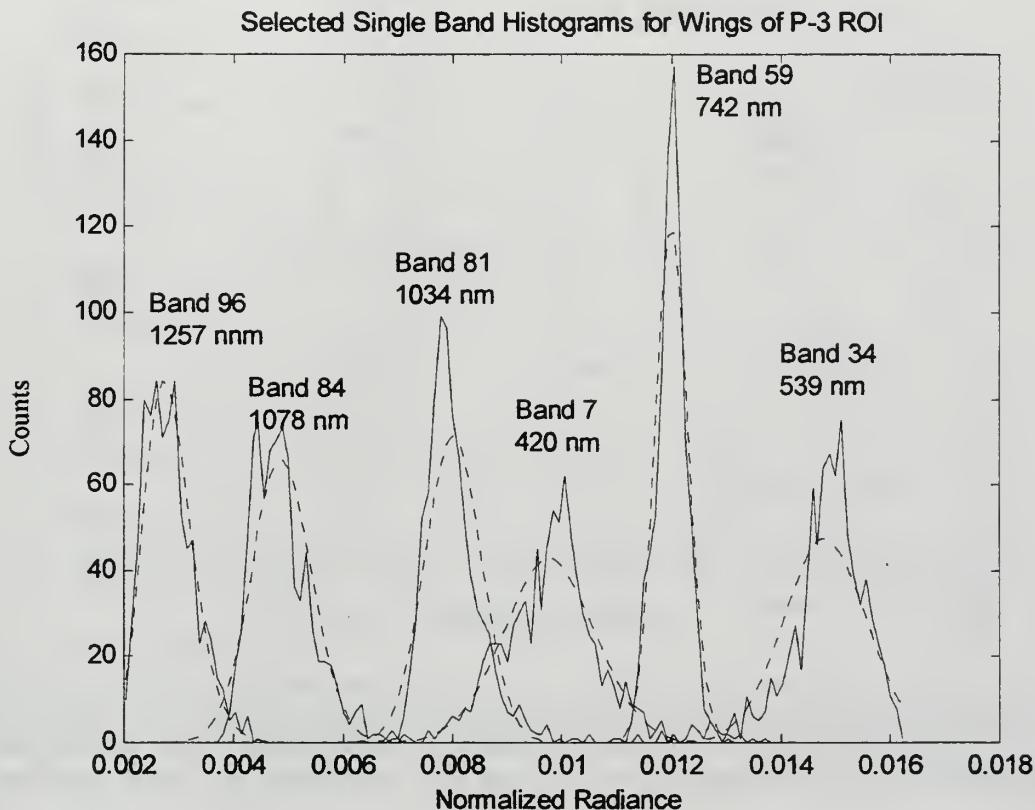
Figure 4.20. Selected single band histograms for Wings of P-3 ROI with dist. fits.

For this figure, the same bands were used as a direct comparison to the plots of the entire P-3 ROI. Bands 7 and 34 were best modeled by the normal distribution while the other four bands were best modeled by the lognormal distribution. The histograms of the wing data did appear more symmetric than the entire P-3 ROI. Table 4.15 shows the results from the wing data. The χ^2 test was used again in the same manner as the C-130

data. The results show that, like the C-130 data, four out of the six selected bands for the wing data were best modeled by the lognormal distribution.

Band	λ(nm)	Normal		Lognormal		Gamma		Best Params
		χ^2	α	χ^2	α	χ^2	α	
7	420	25.40	.00015	38.48	0	34.16	0	$\mu_n = 0.0098$, $\sigma_n^2 = 7.2 \times 10^{-7}$
34	539	77.05	0	94.61	0	88.58	0	$\mu_n = 0.0148$, $\sigma_n^2 = 3.5 \times 10^{-7}$
59	742	23.74	.00021	18.06	.0026	22.62	.0005	$\mu_{bn} = -4.65$, $\sigma_{bn}^2 = 0.022$
81	1034	70.27	0	39.94	0	47.41	0	$\mu_{bn} = -5.16$, $\sigma_{bn}^2 = 0.0101$
84	1078	55.85	0	27.30	0	32.35	0	$\mu_{bn} = -5.33$, $\sigma_{bn}^2 = 0.0120$
96	1257	39.54	0	16.41	.0053	21.80	.0008	$\mu_{bn} = -5.85$, $\sigma_{bn}^2 = 0.0232$

Table 4.15. Goodness of fit for the various histograms presented in Figure 4.20 for the P-3 Wings ROI. MLE Weibull were not always able to converge, and were therefore omitted.

Figure 4.21 shows the selected single band histograms and distribution fits for the Fence Line ROI. The high peaks of the histograms prevented the statistical distributions from fitting properly, although the lognormal distribution fit Band 68 fairly well. The fits of the single band histograms for this ROI are an interesting contrast to the fits of the total radiance histogram. The statistical distributions used here seemed to fit the total radiance histogram qualitatively better than the single band histograms.

72
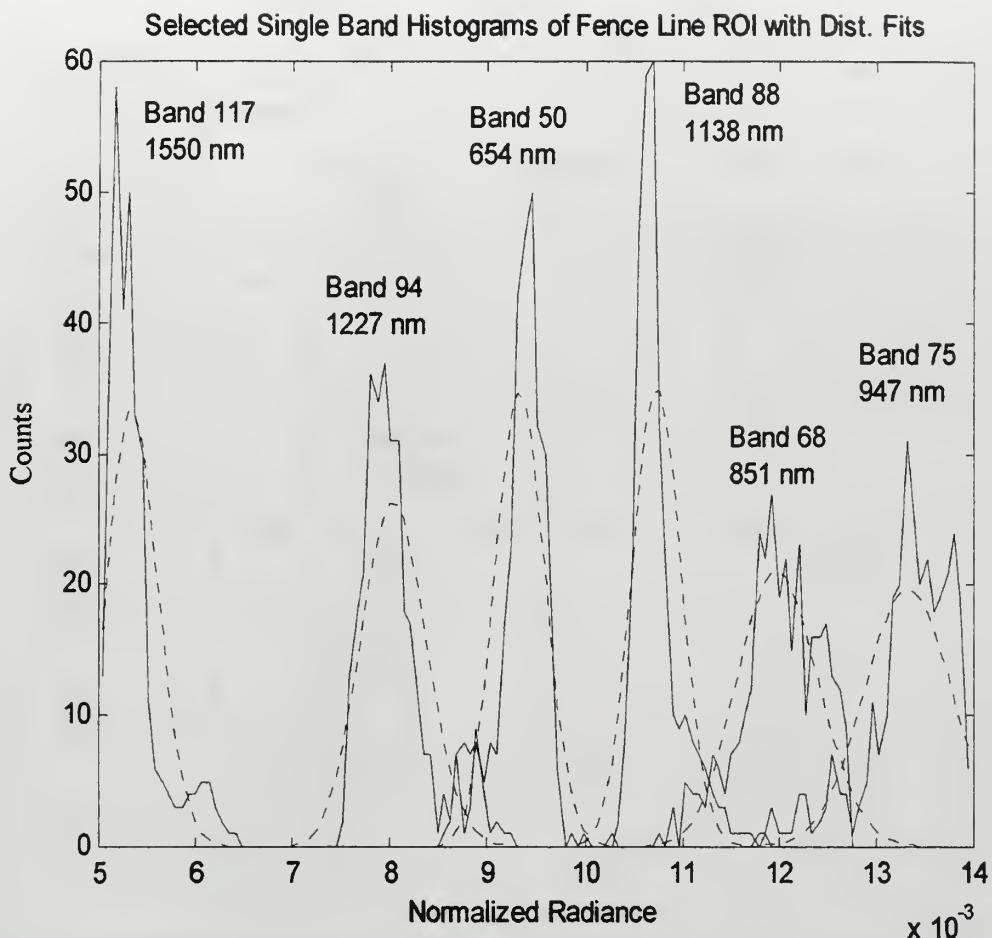
Figure 4.21. Selected single band histograms of Fence Line ROI with dist. fits.

Table 4.16 displays the numerical results of the statistical fits on the histograms from Figure 4.21. Bands 50 and 94 were modeled by the gamma distribution, while the remaining four bands were modeled by the lognormal distribution. Also, Band 68 showed a significance level (α) of .81, which exceeded the other significance levels by far. Again, as was the case with the C-130 and P-3 data, no one distribution proved best for all of the normalized single band histograms.

Band	λ(nm)	Normal		Lognormal		Gamma		Best Params
		χ^2	α	χ^2	α	χ^2	α	
50	654	42.89	0	54.00	0	41.68	0	$\alpha_\Gamma = 872$, $\beta_\Gamma = 1.5 \times 10^{-5}$
68	851	5.21	.56	2.11	.81	3.66	.60	$\mu_{ln} = -4.42$, $\sigma_{ln}^2 = 0.0007$
75	947	99.59	0	79.78	0	82.39	0	$\mu_{ln} = -4.53$, $\sigma_{ln}^2 = 0.0005$
88	1138	31.70	0	25.29	.00016	26.23	.00017	$\mu_{ln} = -4.68$, $\sigma_{ln}^2 = 0.0014$
94	1227	57.13	0	53.52	0	52.53	0	$\alpha_\Gamma = 570$, $\beta_\Gamma = 1.4 \times 10^{-5}$
117	1550	73.94	0	61.81	0	63.45	0	$\mu_{ln} = -5.23$, $\sigma_{ln}^2 = 0.0024$

Table 4.16. Goodness of fit for the various histograms presented in Figure 4.21 from the Fence Line ROI. MLE Weibull were not always able to converge, and were therefore omitted.

Figure 4.22 shows the single band histograms and distribution fits for the first background ROI. The histograms for five of the six bands appear very symmetric and appear to be modeled effectively the statistical distributions. Bands 33 and 65 were best fit to the lognormal distribution, Band 88 was best fit to the gamma, and the remaining bands were best fit to the normal distribution. The distinct, high peaks of the Bands 1, 33 and 65 were not adequately modeled, which led to high values of the chi-squared statistic. The results of the chi-squared test and values of the distribution fits will be shown in the following table.
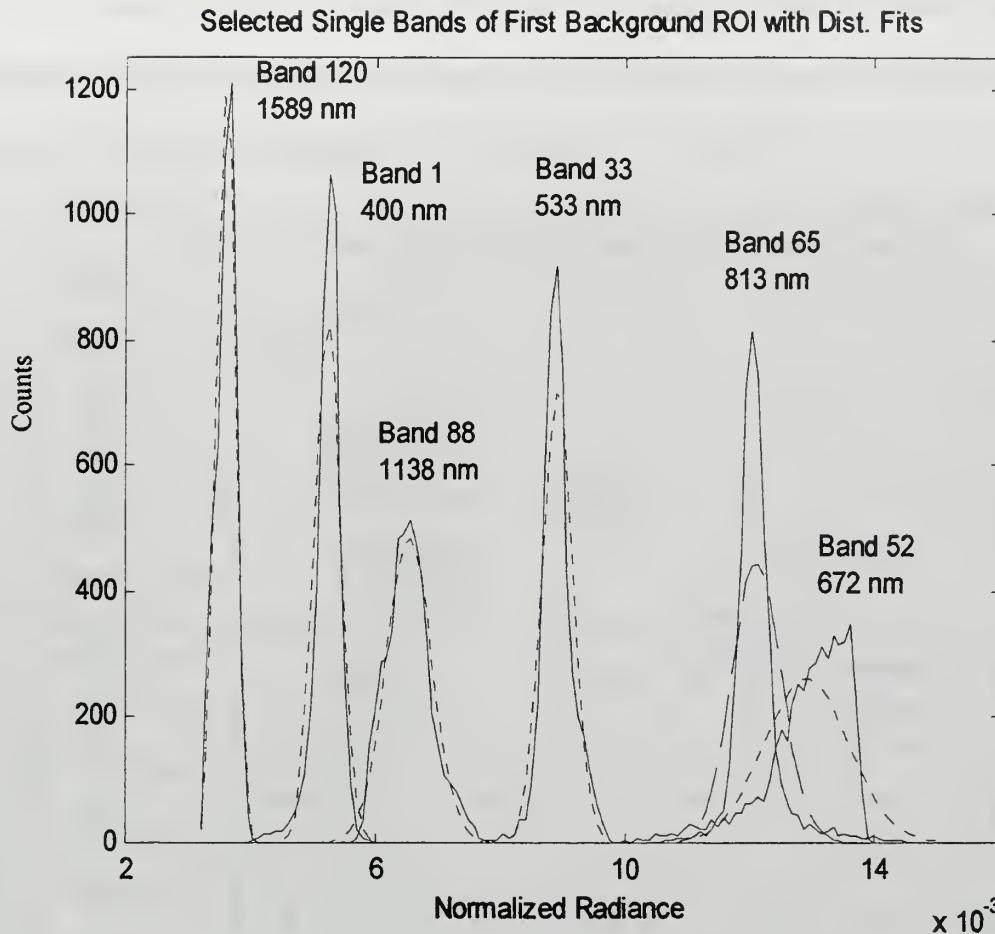
Figure 4.22. Single band histograms of First Background ROI with dist. fits.

Table 4.17 shows the results from the chi-squared goodness of fit test for the first

background ROI. Band 88 was fit best by the gamma distribution with a chi-squared

value of 24.0 and a significance value of .0002. The other bands, which were fit by both

the normal and lognormal, had no chi-squared value less than 178.8 and had a

significance level of zero. This seems in contrast to the visual depiction of the

distribution fits in Figure 4.22 where the distributions appear to fit the normalized single

band histograms, especially in the case of Band 120. However, the chi-squared test

heavily penalizes observations where none are expected, i.e. the tails of the distribution.

Band	λ(nm)	Normal		Lognormal		Gamma		Best Params
		χ^2	α	χ^2	α	χ^2	α	
1	400	421.3	0	541.9	0	498.0	0	$\mu_n = .0052,$ $\sigma_n^2 = 4.6 \times 10^{-8}$
33	533	285.3	0	260	0	267.7	0	$\mu_{ln} = -4.72,$ $\sigma_{ln}^2 = 0.0008$
52	672	823.4	0	1197	0	949.7	0	$\mu_n = .0129,$ $\sigma_n^2 = 4.7 \times 10^{-7}$
65	813	1534	0	1356	0	1420	0	$\mu_{ln} = -4.41,$ $\sigma_{ln}^2 = 0.0011$
88	1138	31.6	0	24.5	.00019	24.0	.0002	$\alpha_\Gamma = 322,$ $\beta_\Gamma = 2.0 \times 10^{-5}$
120	1589	178.8	0	211.7	0	200.4	0	$\mu_n = .0036,$ $\sigma_n^2 = 2.1 \times 10^{-8}$

Table 4.17. Goodness of fit for the various histograms presented in Figure 4.22 from the First Background ROI. MLE Weibull were not always able to converge, and were therefore omitted.

The last normalized single band histograms that will be investigated for the

October over-flight are from the second background ROI. The same bands that were

used for the first background ROI were also used with the second background ROI for

direct comparison between the two. Figure 4.23 shows the normalized single band

histograms from the second background ROI and the corresponding best fits of the

statistical distributions. Visually, the histograms appear symmetric and similar in shape

76

to the first background single band histograms. The statistical distributions do not seem to fit as well as the first background, and the numerical analysis of the fits will be shown in a table following Figure 4.23.
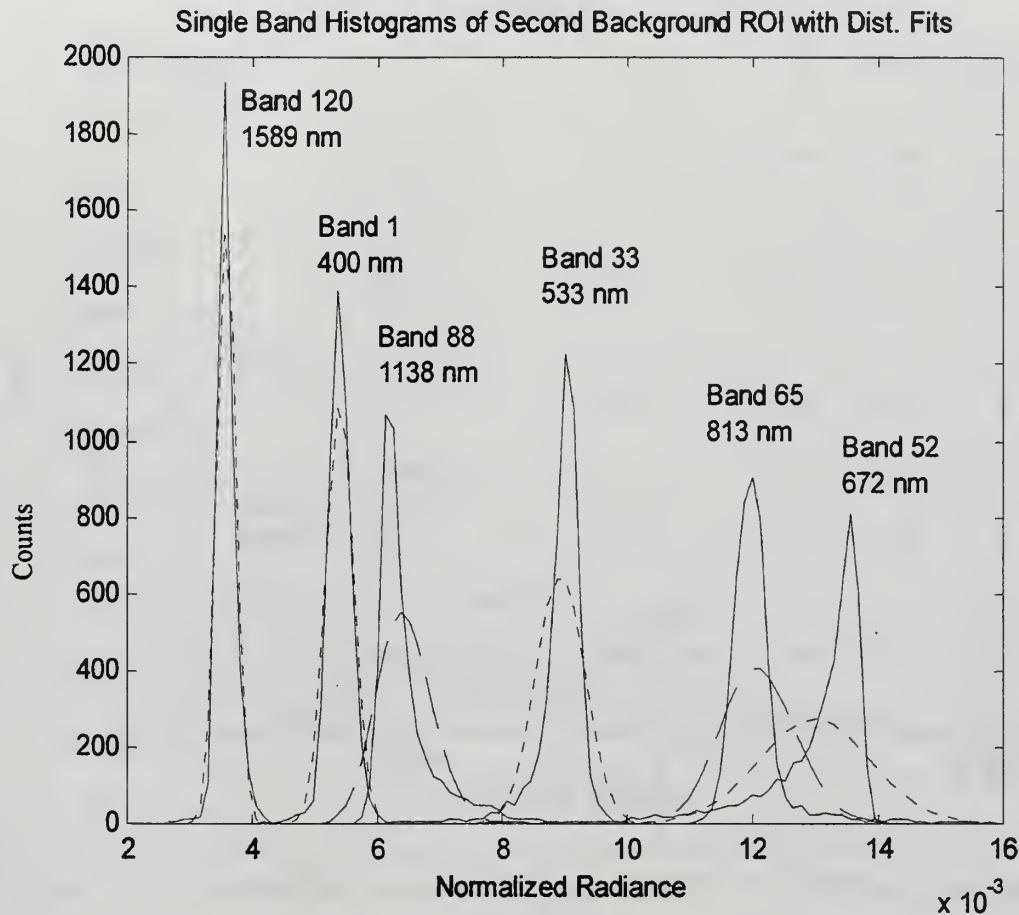

Figure 4.23. Single band histograms of Second Background ROI with dist. fits.

Table 4.18 displays the numerical data from the application of the chi-squared goodness of fit test on the second background ROI histograms. None of the distributions were an adequate fit for the histograms. The chi-squared values were all above 151 and the significance values were all zero. Bands 1, 65, 88, and 120 were approximated by the lognormal distribution and Bands 33 and 52 were approximated by the normal

77

distribution. Again, while in some instances it seemed the statistical distributions would be a good fit qualitatively, the chi-squared test results showed that the fits were not significant.

		Normal		Lognormal		Gamma		
Band	λ(nm)	χ^2	α	χ^2	α	χ^2	α	Best Params
1	400	228.7	0	151.2	0	167.8	0	$\mu_{ln} = -5.22,$ $\sigma_{ln}^2 = .0016$
33	533	1463	0	1822	0	1692	0	$\mu_n = .0089,$ $\sigma_n^2 = 1.4 \times 10^{-7}$
52	672	1416	0	2732	0	2503	0	$\mu_n = .0130,$ $\sigma_n^2 = 8.0 \times 10^{-7}$
65	813	3577	0	2764	0	3016	0	$\mu_{ln} = -4.42,$ $\sigma_{ln}^2 = 0.0025$
88	1138	2329	0	1882	0	2028	0	$\mu_{ln} = -5.05,$ $\sigma_{ln}^2 = .0048$
120	1589	380.0	0	335.6	0	346.8	0	$\mu_{ln} = -5.64,$ $\sigma_{ln}^2 = .0019$

Table 4.18. Goodness of fit for the various histograms presented in Figure 4.23 from the Second Background ROI. MLE Weibull were not always able to converge, and were therefore omitted.

The final statistical analysis that will be accomplished in this section will be to examine the normalized and unnormalized radiance single band histograms and distribution fits from the June over-flight data. The first region that is investigated is the C-130 ROI and Figure 4.24 displays the unnormalized single band histograms with

distribution fits. As a comparison, the normalized single radiance histograms and

distribution fits are plotted in Figure 4.25.
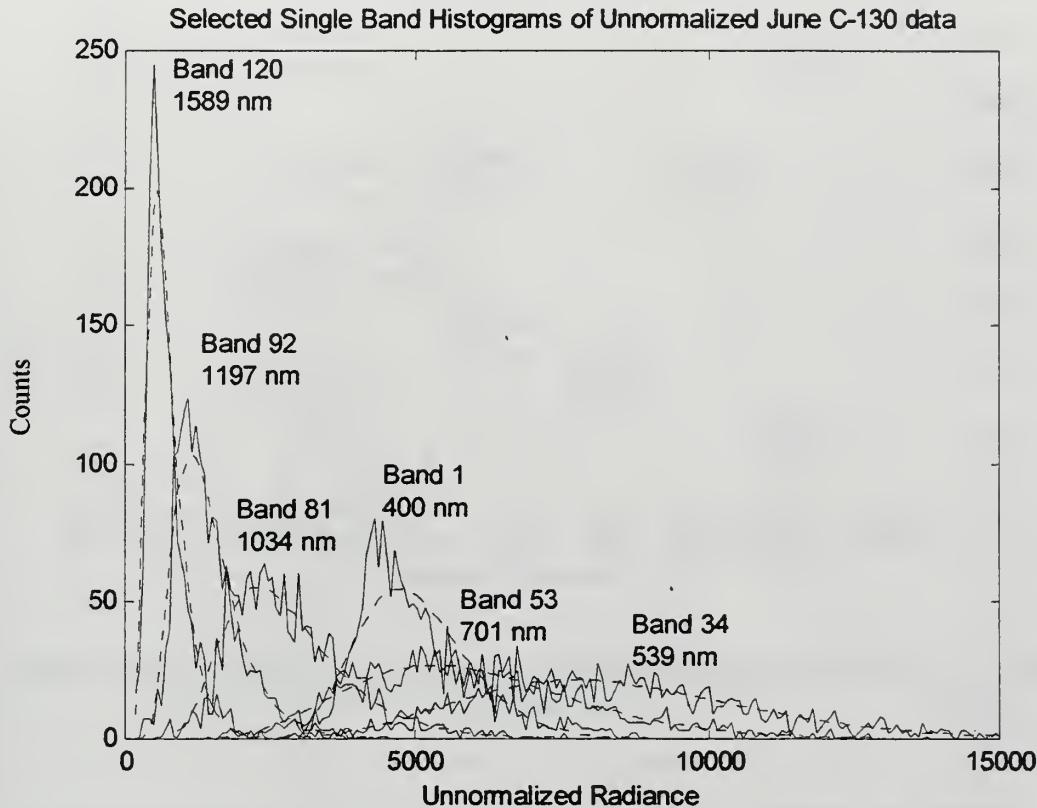
Figure 4.24. Unnormalized single band histograms of C-130 ROI with best fitting
distributions.

Bands 34 and 53 were modeled best by the gamma and had low chi-squared

values of 3.53 and 8.70 respectively. The other four bands were fit best the lognormal

distribution. The normal distribution produced chi-squared values in excess of 41 and did

not model any of the histograms with any degree of statistical significance.
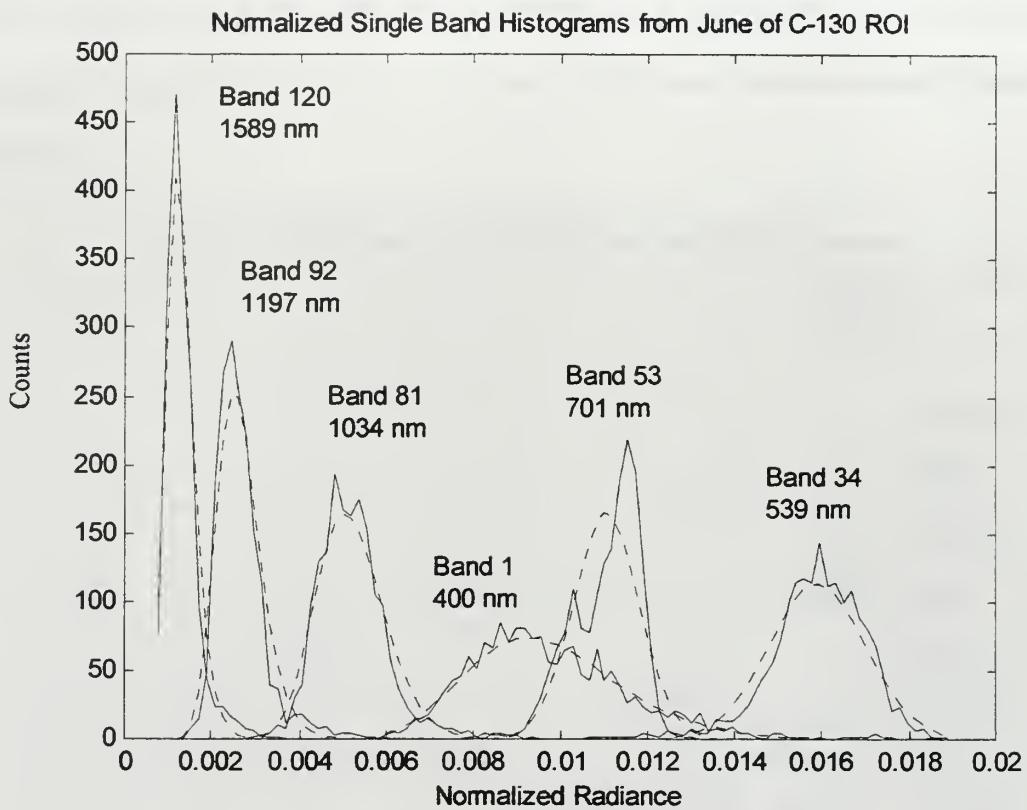
Figure 4.25. Normalized Single Band Histograms of C-130 ROI with Best Fitting Distributions.

Visually, the statistical distributions appear to model the histograms fairly well. Bands 34 and 53 were fit best by the normal distribution while Bands 1, 81, 92, and 120 were fit best by the lognormal. The only significant fit, however, as the chi-squared test revealed, was Band 1, which had a χ^2 value of 10.28 and a significance level of .0677.

The unnormalized single band histograms with distribution fits for the P-3 Wing ROI are plotted in Figure 4.26. As with many of the single band histograms presented in this thesis, the lognormal distribution was the best fit for four of the six bands while Band 7 was best fit by the normal and Band 34 was best fit by the gamma. Results of the χ^2

test showed that Band 7 had the lowest chi-squared value of 6.24 and Band 34 had a chi-squared value of 11.48. The other four bands had chi-squared values in excess of 21.
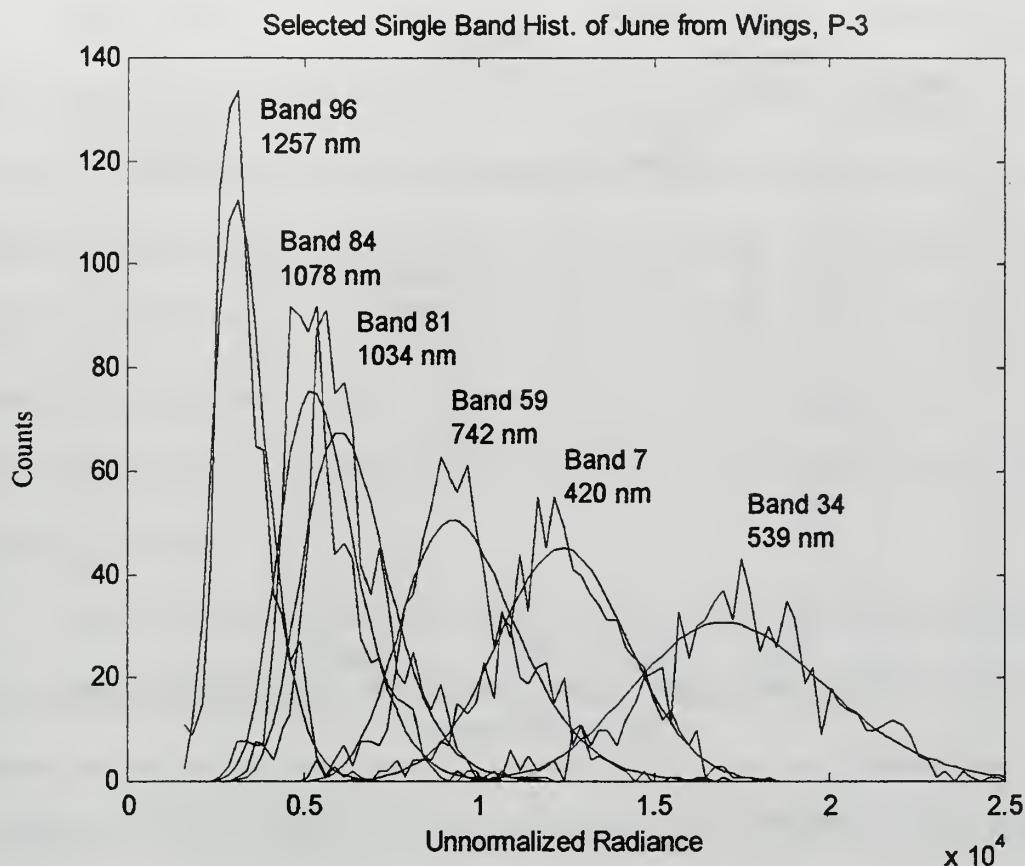

Figure 4.26. Unnormalized single band histograms of P-3 Wings ROI with dist. fits.

Figure 4.27 shows the normalized radiance single band histograms of the P-3 Wings ROI and the corresponding best statistical distribution fits. Except for Band 34, which was fit best by the gamma distribution, all of the histograms were fit best by the lognormal distribution. After computing the χ^2 for each of the bands and distribution fits, Band 96 had the lowest value at 8.437 with a significance level of .1337. The chi-squared values for the normalized P-3 Wing data from the June over-flight were

81

comparable to the values of the normalized October over-flight data.

**Single Band Histograms from June of Wings, P3**
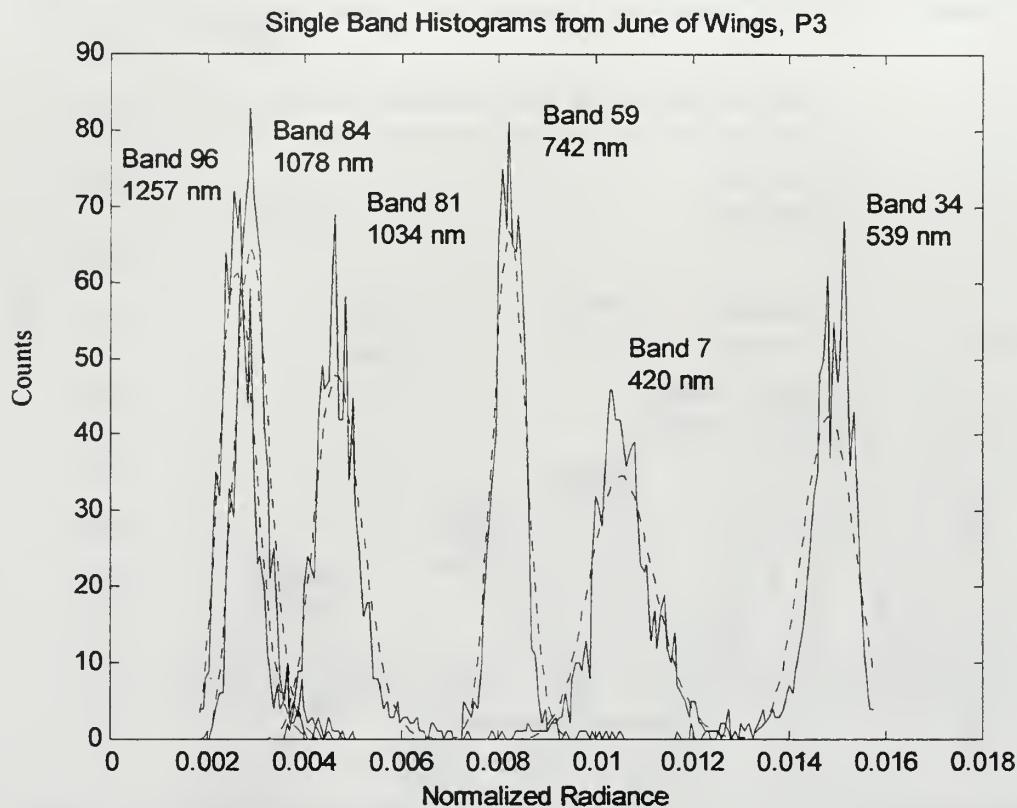
Figure 4.27. Normalized single band histograms of P-3 Wings ROI with dist. fits.

This chapter has detailed an extensive statistical analysis on the spectral data from Chapter III. The total radiance, single band unnormalized and single band normalized histograms were modeled by the four MLEs discussed previously in an effort to discover the true nature of hyperspectral data distributions. Chapter V discusses the findings and implications of the work accomplished in Chapter IV.

V.   DISCUSSION

Chapters III and IV of this thesis have looked at several different aspects of the Davis Monthan (DM) data in an effort to explore the prevalent use of normal, Gaussian statistics to model data distributions in hyperspectral scenes. While this thesis explored thousands of individual pixels, each across 210 spectral bands in two separate over-flights of the DM scene, it was still only a fraction of the data available for research. However, it is the hope of the author that the findings in this analysis will give rise to new avenues of research into hyperspectral imagery, specifically in the area of spectral imagery classifiers.

After examining the total radiance, normalized single band and unnormalized single band spectral radiance distributions for each of the eight regions of interest, it has become evident that normal statistics do not adequately model data distributions in hyperspectral scenes. While some of the data distributions appeared symmetric, when a normal Maximum Likelihood Estimator (MLE) was fit to the data distribution and a goodness of fit test was performed, it became obvious that the data did not follow a normal distribution. For example, after applying the χ^2 goodness of fit test to the total radiance histogram of the C-130 ROI from the October over-flight, the statistical significance of the normal distribution was shown to be exactly zero (to floating point precision). Also, using the χ^2 test for the single band normalized data from each ROI revealed an average chi-squared value across the selected bands for the normal distribution fit of 386.83 with a significance level of zero. And, the single band unnormalized data was an even worse fit to the normal distribution, with the average chi-

squared value across the selected bands being 550.34 and significance of zero. The majority of spectral data processing schemes (matched filter, pixel unmixing, pixel classification) currently use multi-variate normal statistics, and based on the above data, should be investigating the use of different statistical distributions.

While the normal distribution was not a good fit for the DM data as a whole, it was able to model some of the individual bands within each ROI. For example, the chi-squared value for the normalized Band 68 data from the Fence Line ROI of the October over-flight was 5.21 with a significance level of .56. Also, the chi-squared value for the normalized Band 7 data from the P-3 Wings ROI of the June over-flight was 6.24 with a significance level of .43. However, typically just one band from the selected bands of each ROI was fit well by the normal distribution, which, by itself, is not significant, however, it is likely that other bands would have similar success since the one band is representative of one type of band histogram in that ROI.

It has been found that an effective tool for examining the DM data has been the χ^2 goodness of fit test. It was used to show how well each of the MLEs fit the corresponding histograms of the DM data. For every region of interest studied here, the lognormal distribution had the lowest or best chi-squared value for four out of the six bands that were selected from each ROI. The results from the chi-squared tests were again, further evidence that the lognormal distribution was a reasonable model for the DM spectral data. Table 5.1 shows the average chi-squared values, computed using 5 degrees of freedom, for each of the distributions.

84

Distribution	Average chi-squared values		
	Total Radiance	Single Unnormalized	Single Normalized
Normal	640.61	235.35	468.59
Lognormal	344.39	110.64	319.31
Gamma	551.10	187.34	413.78

Table 5.1. Statistical Distributions with Average Chi-squared Values.

The lognormal and gamma distributions, as stated previously in Chapter IV, always have right tails which enabled them to model the right-tailed data histograms well, as was the case with the C-130 ROI and most of the total radiance and unnormalized data, while the most symmetric single band histograms were modeled best by the normal distribution, which by its nature is symmetric. However, none of the distributions were able to model left tails effectively, which were evident in a few of the background ROI single band histograms. Investigation into distributions that can model left-tailed histograms is needed to advance the research done in this thesis. The data used in this study had a majority of right-tailed data distributions, which is the reason the lognormal and gamma distributions were shown to be the best models for this research. While a typically small percentage of events will occur in the tails of the distribution, it is precisely these events which will lead to false alarms (long right tails in background distributions) and misses (long left tails in target distributions). The results of the χ^2 test, while showing that the statistical significance of the three distributions was low, did indicate that the lognormal, and gamma distribution in some instances, was able to perform far better than the normal distribution in the modeling of the distribution tails.

The comparison between the June and October HYDICE over-flight data was crucial in showing the stability of the results presented here. Overall, the results were very similar, demonstrating the viability of the findings in this thesis. The following four figures show a direct comparison between the C-130 and P-3 Wing ROIs from both over-flights. Figure 5.1 shows the C-130 ROIs and the distribution fit of the lognormal and Figure 5.2 shows the C-130 ROIs and the distribution fit of the normal. Table 5.2 will list the goodness-of-fit data for these plots.
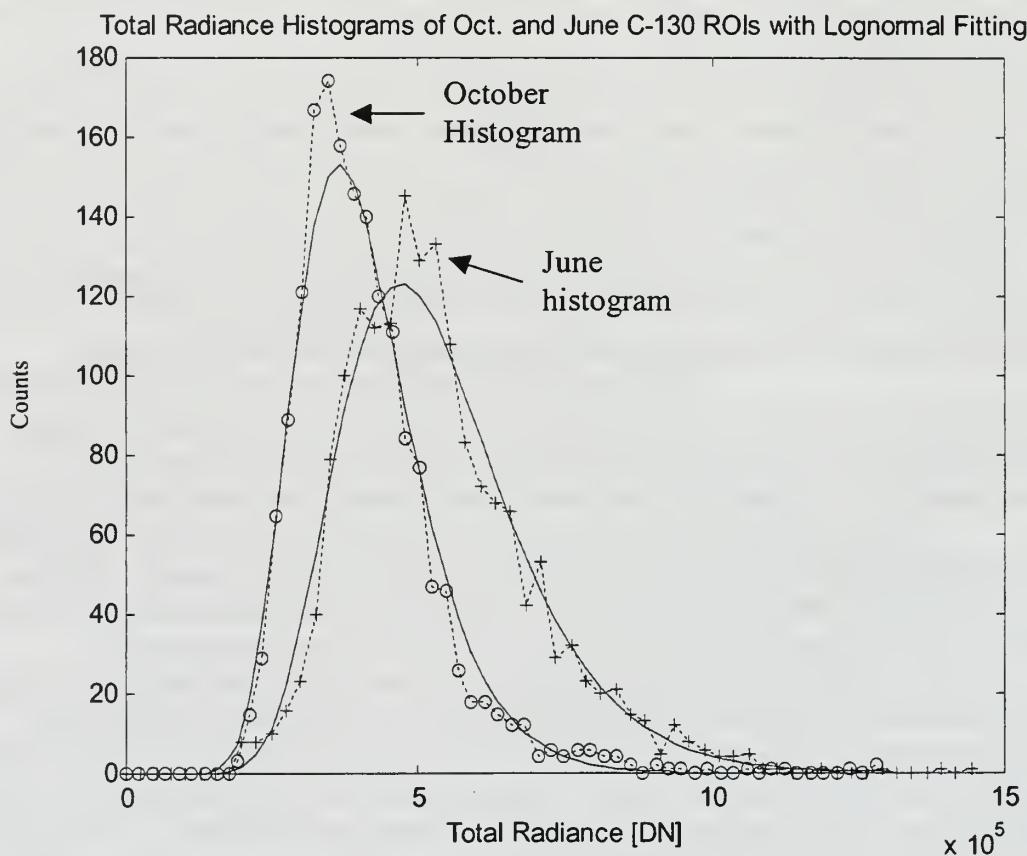

Figure 5.1. Total radiance histograms of C-130 ROIs from both over-flights with Lognormal distribution fits. The solid curves represent the lognormal distribution fit.
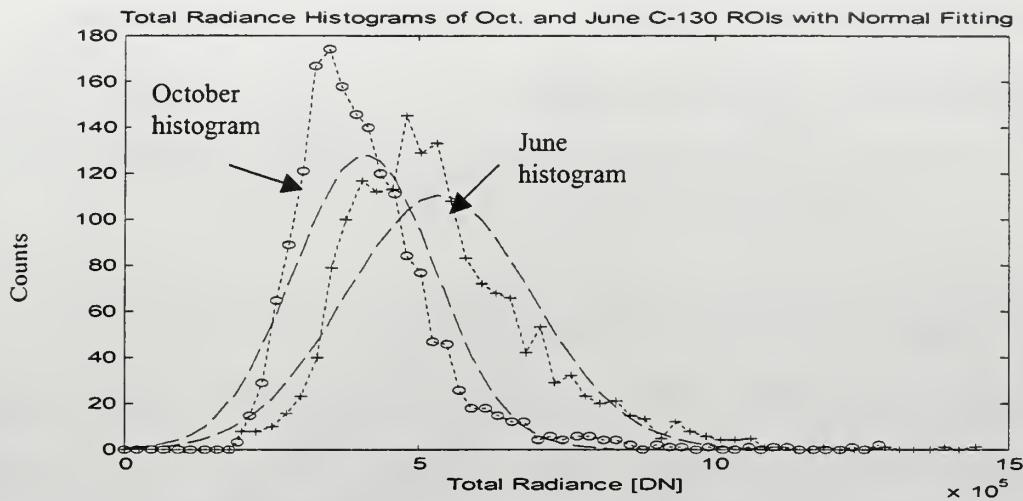
Figure 5.2. Total radiance histograms of C-130 ROIs from both over-flights with Normal distribution fits. The dashed curves represent the normal distribution fit.

The plots in Figures 5.1 and 5.2 show the similarity between the two over-flights.

The histograms of the data are both left skewed with right tails and the lognormal distribution was the best fit for both while the normal fit was insignificant. Table 5.2 shows the goodness-of-fit parameters for the lognormal and normal distributions for these two ROI.

Over-flight/ROI	Normal		Lognormal	
	χ^2	α	χ^2	α
June/C-130	71.66	0	8.76	0.23
Oct./C-130	65.34	0	11.03	0.05

Table 5.2. Goodness-of-fit data for C-130 ROI.

Another example of similar distributions and fits was the P-3 Wing ROI data from both over-flights. Figures 5.3 and 5.4 show the histogram and distribution fits for the P-3 Wing ROI from the October and June over-flights. Both histograms are slightly left skewed with short, right tails and again, the lognormal distribution is the best model in both plots while the normal is the worst model. While these four figures did not

87

represent all of the characteristics of the two over-flights, they do show that there is

inherent stability in the results presented here.
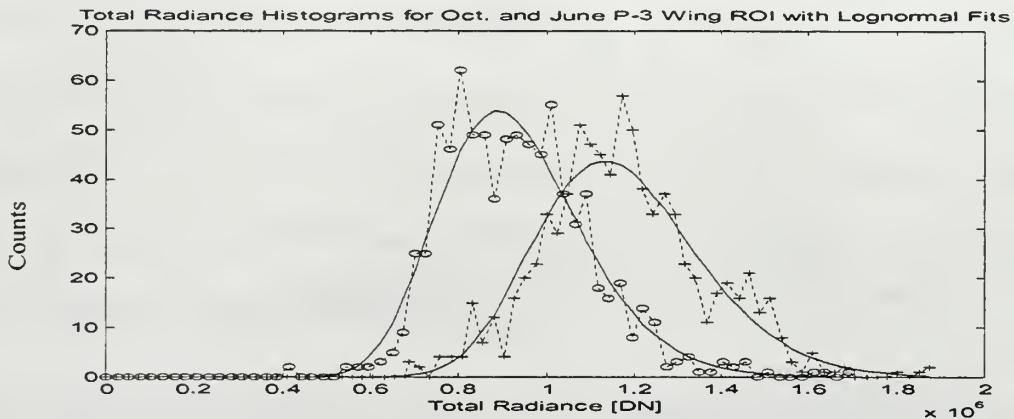
Figure 5.3. Total radiance histograms of P-3 Wing ROIs from both over-flights with Lognormal distribution fits. The solid curves represent the lognormal distribution fit.
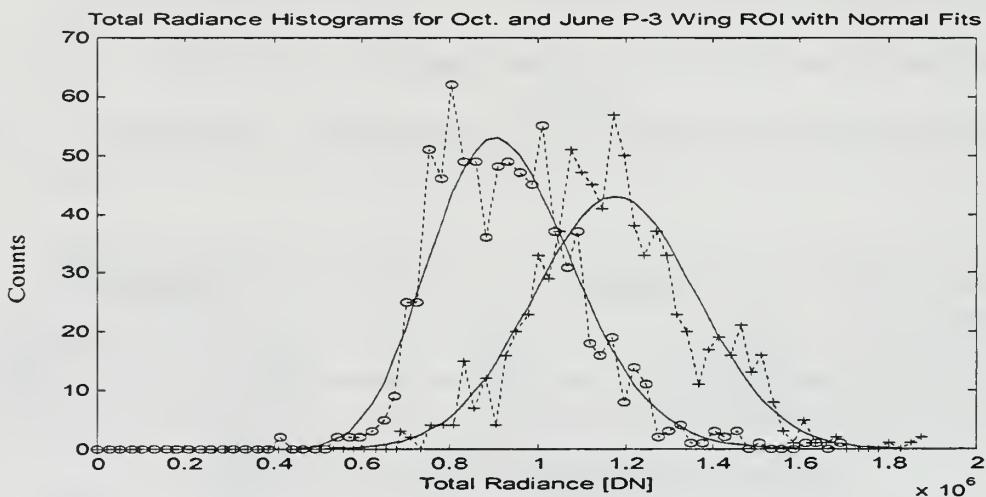

Figure 5.4. Total radiance histograms of P-3 Wing ROIs from both over-flights with Normal distribution fits. The solid curves represent the normal distribution fit.

Table 5.3 displays the goodness-of-fit parameters for the P-3 Wing ROI

distribution fits. While the fit is better for the June data, this table still shows that that

lognormal distribution fit better than the normal distribution for the data from both over-

flights.

Over-flight/ROI	Normal		Lognormal	
	χ^2	α	χ^2	α
June/P-3 Wing	11.74	0.048	5.31	0.65
Oct./P-3 Wing	159.32	0	106.04	0

Table 5.3. Goodness-of-fit data for P-3 Wing ROI.

A comparison of the mean and variance values of the two over-flights also gives a good indication of the similarities between the two. Table 5.4 shows the mean and variance values for both the lognormal and normal distributions from the C-130 and P-3 Wing ROI total radiance data from both over-flights. For both sets of data, the mean and variance values for the normal distribution were high and varied a great deal. Also for both over-flights, the mean and variance values of the lognormal distribution for both ROIs were low and fairly constant. Also, while the lognormal parameters were stable, the normal distribution parameters were unstable and less accurate. The implication of the similarities between the October and June HYDICE over-flights cannot be understated. The lognormal and normal distributions fit similar for both over-flights and the nature of the distribution parameters was comparable as well, which means that the research conducted in this thesis is viable for use with other hyperspectral scenes and under different environmental conditions.

Over-flight/ROI	Normal			Lognormal		
	Mean[x]	Variance[x]	$\frac{\sqrt{\sigma^2}}{\mu}$	Mean[x]	Variance[x]	$\frac{\sqrt{\sigma^2}}{\mu}$
June/C-130	5.3×10^5	2.5×10^{10}	.2983	13.14	0.084	.0221
Oct./C-130	7.7×10^5	8.7×10^{10}	.3830	13.49	0.109	.0245
June/P-3 Wing	1.2×10^5	3.5×10^{10}	1.559	13.97	0.025	.0113
Oct./P-3 Wing	7.1×10^5	8.1×10^{10}	.4009	13.46	0.022	.0110

Table 5.4. Mean and Variance Values for C-130 and P-3 Wing ROI from Both Over-flights.

89

The different statistics used in this research gave insight into the nature of the Davis Monthan spectral data. When the total radiance histograms were fit by the MLE distributions beginning with the C-130 ROI in Figure 4.1, a clear pattern began emerging. The lognormal distribution was discovered to be the best fit for each of the total radiance histograms in both HYDICE over-flights, a pattern that would continue with the majority of single band histograms being fit to the lognormal as well.

In order to examine all aspects of the hyperspectral data, both unnormalized and normalized single band histograms were used in the research for this thesis. The general trend seen for the normalized data was that bands exhibiting the most symmetric histograms were from the visible portion of the spectrum, bands that were at the edge of the visible were skewed with left tails, and the bands in the infrared portion of the spectrum were skewed with right tails. When the MLE distributions were fit to both of these types of data, the lognormal distribution was found to model most individual band histograms for both the normalized and unnormalized data, further indicating that this distribution could be an effective model for this spectral data. The quality of fit of the lognormal distribution was statistically insignificant for many of the bands, but even if the fit for these two types of data was not ideal, it still provided a significant improvement over the normal distribution fit.

At this point, it is worthwhile to ask why the distribution of the total radiance might be lognormal in hyperspectral imagery. A lognormal distribution results in a random variable that is normal distributed when its natural logarithm is taken. If it is assumed that the total radiance is the product of several random variables as

90

$$L_{tot} = L_{solar} \cdot T_{down} \cdot T_{shade} \cdots R_{surf} \cdot R_{shape} \cdot T_{up},$$ (5.1)

where the various terms in Equation 5.1 correspond to solar radiance, downward and upward atmospheric transmission, shape, shading, and spectral reflectance. Taking the natural logarithm,

$$\ln L_{tot} = \ln L_{solar} + \cdots + \ln T_{up}.$$ (5.2)

Each of the terms on the right hand side of Equation 5.2 is itself a random variable, and their sum is expected to be normal distributed by the central limit theorem (Devore, 1995). In this sense, the variations can to a large extent be considered as

$$L = \underline{\underline{R}} \cdot L_{solar} + n,$$ (5.3)

where $\underline{\underline{R}}$ is a random operator (typically a diagonal matrix) corresponding to the atmospheric and target effects and n is a random noise vector added at the detection process.

The use of lognormal statistics to describe hyperspectral data can be used in a number of applications. The spectral analysis techniques of matched filters and pixel unmixing with noise have regularly assumed Multivariate Normal statistics. By taking the logarithm of the spectral data, these same tools can be applied with presumably better success. Also, the use of the K-means pixel classification algorithm could be improved by directly implementing lognormal, instead of normal statistics when classifying pixels. Figure 5.5 shows a pictorial representation of the K-means algorithm, and how it could be improved with lognormal statistics.
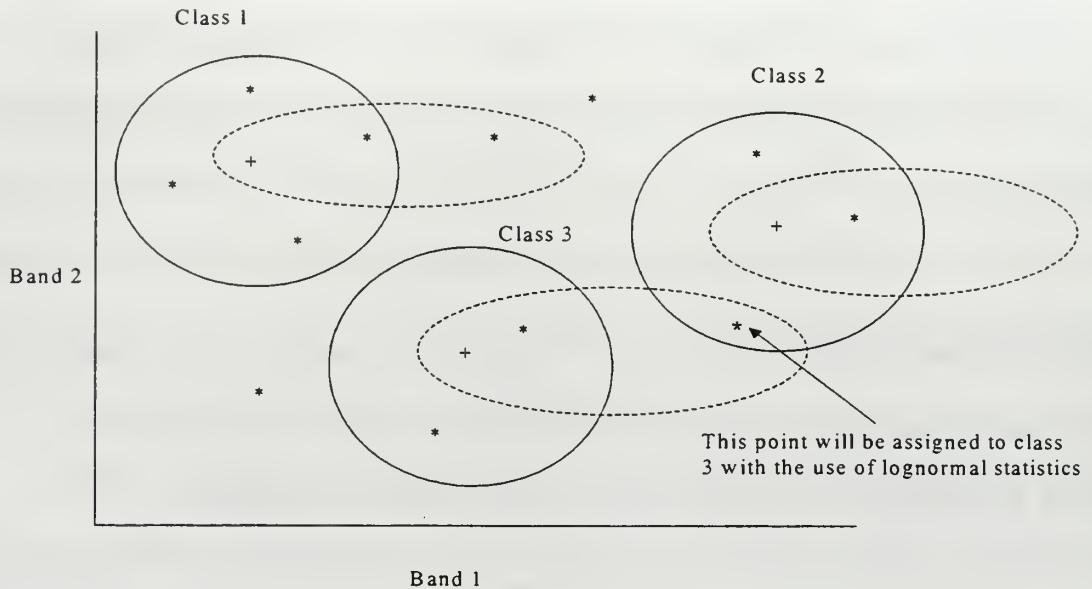
Figure 5.5. K-means algorithm representation. The circles represent multi-variate normal statistical classification while the ovals represent lognormal statistics. The stars represent individual pixels. Pixels will be assigned to different classes depending on the type of distribution used.

One interesting point that was discovered when looking at the total radiance histograms was the fact that the P-3 ROI had a bimodal distribution. The bimodal nature of this ROI, which was found to be primarily caused by the fuselage of the P-3 aircraft, was clearly not going to be modeled by a unimodal distribution. This characteristic was also seen in both the unnormalized and normalized single band histograms of the P-3 data. The implication of the bimodal nature of this ROI is that unimodal statistics, as were the distributions that were investigated here, will not be able to model this type of data. Multi-modal statistics that can account for this nature will have to be used in order to be able to handle these types of distributions.

As was previously discussed in Chapter IV, the Weibull distribution did not prove to be a useful statistical tool in this thesis. The MATLAB program had difficulty implementing it and often could not provide accurate results. As this distribution has been used previously in RADAR applications with some success (Stein, 1999), it could be worth revisiting in the future since it has a left tail, to see if it can be applied successfully to spectral data with left-tailed distributions. Also, the use of other statistical models with left tails, such as the Beta distribution, could give promising results in the analysis of hyperspectral data.

The inherent problem of poor quality of fit and statistical significance, even for the lognormal distribution, needs to be addressed in future work. While the lognormal provided a significant improvement over the normal distribution, it often was not able to achieve a statistically significant fit of the various data histograms. A cursory look into different methods of parameter estimation led to trying the method of minimization of total chi-squared. This method searches for the parameters (μ, σ^2 for lognormal) that minimizes the value of the χ^2 statistic in Equation 4.5. While this fit is not the maximum likelihood fit, it is the one that has the highest significance value under the χ^2 test. Applying this method to the C-130 ROI from the October over-flight, the χ^2 value of the lognormal dropped when compared to the value from the MLE. Overall, the significance level increased for this distribution from .05 to .43. The gamma and normal distributions also improved but the significance level was less than .001 for each. Table 5.5 shows the results of the application of this method.

	MLE		Min. of χ^2	
	χ^2	α	χ^2	α
Lognormal	11.03	.050	4.85	.43
Gamma	26.29	0	20.55	9×10^{-4}
Normal	113.41	0	31.0	9×10^{-6}

Table 5.5. Comparison between MLE and Minimization of χ^2.

In essence, this method improved the overall fit of the statistical distributions. This interesting finding gives a good basis for continuing work in this area of research.

VI.   CONCLUSIONS

In this thesis, Chapter I introduced the area of hyperspectral imaging that was investigated and presented the goals of this study. Chapter II gave a brief overview into the background of hyperspectral imaging and described various HSI analysis techniques. Chapter III gave a detailed description of the scene from which hyperspectral data was extracted and studied and also presented total radiance, unnormalized radiance, and normalized radiance histograms of this data. Chapter IV gave exacting detail of the statistical analysis conducted on the hyperspectral data and provided numerous plots and numerical data. Chapter V was a discussion of the findings and implications of the work carried out in this thesis.

From the findings in this thesis, the lognormal distribution should be considered for use in the area of hyperspectral imagery analysis. Use of the lognormal with other data sets from various hyperspectral scenes will further validate the findings that have been introduced in this study. The lognormal distribution has been shown to be an effective model of total radiance, unnormalized radiance in single bands, and, to a lesser extent, normalized radiance in single band distributions for the Davis Monthan data. The normal distribution, which has been used extensively to describe the nature of spectral signatures of targets in spectral scenes, was shown to be an inadequate model for the distributions of spectral data. Also, there is a need to explore multi-variate distributions to see if the results can truly be extended to HSI processing.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX: MATLAB AND IDL SOURCE CODE

```
b = normalize('fencelineroi.txt',rad) ;can change roi.txt file to view data
from
                                ;various rois
                                ;rad is a variable that is total radiance
                                ;for each pixel
                                ;b is the normalized radiance as a
                                ;function of wavelength
                                ;for each pixel


desired_bands=[120]  ;selected bands from roi file

range = max(rad)-min(rad)

  nbins = 50

  totalradhist = histogram(rad,binsize = range/nbins)
  ;this computes a histogram of total radiance across all bands
  ;divided into a set number of bins

nb=n_elements(desired_bands)

x_tot = min(rad)+indgen(n_elements(totalradhist))*range/nbins

x_axis_values = fltarr(76,nb)

gauss_values = fltarr(76,nb)

histograms = intarr(76,nb)

for m = 0,n_elements(desired_bands)-1 do begin

  band = desired_bands[m]

  range = max(b[band,*])-min(b[band,*])

  nbins = 30

hist = histogram(b[band,*],binsize = range/nbins)  ;computes a histogram of
                                              ;normalized data
                                              ;with a selected bin size
x_axis = min(b[band,*])+indgen(n_elements(hist))*range/nbins
                                ;generates a
                                ;vector which is same length as
                                ;hist and covers the range from
                                ;min to max

  ;gauss = gaussfit(x_axis,hist,coeffs,nterms=3) ;fits a gaussian to our data

  mle_gauss,b[band,*],sigma2_hat,mu_hat

  for k = 0,n_elements(hist)-1 do histograms[k,m]=hist[k]
```

97

```
  for l = 0,n_elements(hist)-1 do x_axis_values[l,m]=x_axis[l]

  for o = 0,n_elements(hist)-1 do

gauss_values[o,m]=1/sqrt(sigma2_hat*2*!pi)*exp(-(0.5)*(x_axis[o]-

mu_hat)^2/sigma2_hat)

 endfor

end
```

```
%function [chi2,p]= chi2fit(hgram,dist,k,p)
%this is a function that computes the chi^2 parameter and the associated
%probability given a histogram and a distribution
%inputs:
%  hgram - histogram
%  dist - distribution function
%  m - number of bins to analyze
%  p - number of parameters estimated

function [chi2,p]= chi2fit(hgram,dist,m,p)
hgram=hgram(:);dist=dist(:);
dist=round(dist*sum(hgram)/sum(dist));%get the proper normalization

N=sum(hgram);%total number of observations
DOF=m-p-1;
spacing = round(N/m);
count = spacing:spacing:N;
count(end)=N; %these set the number of counts up so there are the right number.
               %the last count must be N, so all trials are accounted for.
phi_obs=cumsum(hgram);%observed cumulative distribution
phi_exp=cumsum(dist);%expected cumulative distribution

for k=1:length(count)%find the indices corresponding loosely to these counts
                     %in the expected distribution
    ind(k)=max(find(phi_exp<=count(k)));
end ·

ind=[1 ind];%need the first bin in the hgram

exp=phi_exp(ind(2:end))-phi_exp(ind(1:end-1));
obs=phi_obs(ind(2:end))-phi_obs(ind(1:end-1));

if find(exp==0)
    chi2=NaN;
    p=NaN;
    disp('Not enough bins in the histogram to compute with this many DOF')
    return
end

chi2=sum((exp-obs).^2./exp);
p=chi2cdf(chi2,DOF);
```

```
%this is a procedure that I'm writing to do the individual band calculations
%for statistics.

clear

%Get the normalized data (from IDL)
fid=fopen('fence.dat');
b=fread(fid,[210,inf],'int16');
b=b.';
fclose(fid);

%get the unnormalized data
%fid=fopen('june-c130.dat');
%a=fread(fid,[210,1740],'int16');
%a=a.';
%fclose(fid);

desired=[round(1:6.5:65),80:3:85,92:2:100,120:4:140];%the band numbers

%desired=[10 30 50 85 120];               %that I want to process
desired=desired([5 7 8 11 13 17]);        %I got these by examining
                                          %the mean spectrum

%desired=desired(1);

[hgrams,x]=hist(b(:,desired),300);%compute the column by column histogram

for k=1:length(desired)
    disp(['k = ' int2str(k)])
    data = b(:,desired(k));%pick the data

    [m_n(k),s_n(k),h_n(:,k)]=norm_fit(data,x);disp('normal')
    [m_l(k),s_l(k),h_l(:,k)]=lognorm_fit(data,x);disp('lognormal')
    [a_g(k),b_g(k),h_g(:,k)]=gam_fit(data,x);disp('gamma')
    [a_w(k),b_w(k),h_w(:,k)]=weib_fit(data,x);disp('Weibull')

    [xn(k),pn(k)]=chi2fit(hgrams(:,k),h_n(:,k),8,2);
    [xl(k),pl(k)]=chi2fit(hgrams(:,k),h_l(:,k),8,2);
    [xg(k),pg(k)]=chi2fit(hgrams(:,k),h_g(:,k),8,2);
    [xw(k),pw(k)]=chi2fit(hgrams(:,k),h_w(:,k),8,2);

end
```

```
%total rad proc.m
%this is a procedure which I'm using to do statistical modeling on
%the total radiance data from the DM image.

fid = fopen('c130.dat');%this is something I get from IDL.  It has the
                        %total radiance for each pixel in the roi
rad=fread(fid,inf,'int');
fclose(fid);

%load p3_rad.asc -ascii
%rad=p3_rad(1:end-3);
N=length(rad);%total number of pixels

nb=50;%number of bins to use in the histogram
[hgram,x]=hist(rad,nb);%compute the nb-bin histogram

dx=x(2)-x(1);
hgram=[zeros(size(0:dx:x(1))) hgram];%pad with zeros to the left
x=[0:dx:x(1) x];

[mu_nor,sig2_nor]=mle_gauss(rad);%MLE estimator of the gaussian params
nor_pdf=1/sqrt(2*pi*sig2_nor)*exp(-(1/2).*(x-mu_nor).^2/sig2_nor);
nor_hist=nor_pdf*dx*N;

[mu_ln,sig2_ln]=mle_gauss(log(rad));%MLE estimator of lognormal params
ln_pdf=1./(sqrt(2*pi*sig2_ln)*x).*exp(-(1/2)*(log(x)-mu_ln).^2/sig2_ln);
ln_pdf(isnan(ln_pdf))=zeros(size(find(isnan(ln_pdf))));%log of zero probs.
ln_hist=ln_pdf*dx*N;

phat=weibfit(rad);%MLE Weibull distribution (Stats toolbox)
al_w=phat(1);be_w=phat(2);
we_pdf=weibpdf(x,al_w,be_w);
5we_hist=we_pdf*dx*N;

phat=gamfit(rad);%MLE Gamma distribution (Stats toolbox)
al_g=phat(1);be_g=phat(2);
ga_pdf=gampdf(x,al_g,be_g);
ga_hist=ga_pdf*dx*N;

[a_b,b_b,be_hist]=beta_fit(rad,x(2:end));%MLE Beta Distribution
```

```
%total rad proc.m
%this is a procedure which I'm using to do statistical modeling on
%the total radiance data from the DM image.

fid = fopen('c130.dat');%this is something I spit out from IDL.  It has the
                                                %total radiance for each pixel
in the roi
rad=fread(fid,inf,'float32');
fclose(fid);

N=length(rad);%total number of pixels

nb=100;%number of bins to use in the histogram
[hgram,x]=hist(rad,nb);%compute the nb-bin histogram

data = rad;

save data data nb

dx=x(2)-x(1);
hgram=[zeros(size(0:dx:x(1))) hgram];%pad with zeros to the left
x=[0:dx:x(1) x];

[mu_nor,sig2_nor]=mle_gauss(rad);%MLE estimator of the gaussian params
phat=fminsearch('chi2_norm',[mu_nor sig2_nor]);%estimator of the gaussian
params
mu_nor=phat(1);sig2_nor=phat(2);
nor_pdf=1/sqrt(2*pi*sig2_nor)*exp(-(1/2).*(x-mu_nor).^2/sig2_nor);
nor_hist=nor_pdf*dx*N;

[mu_ln,sig2_ln]=mle_gauss(log(rad));%MLE estimator of lognormal params
phat=fminsearch('chi2_ln',[mu_ln sig2_ln]);%estimator of the gaussian params
mu_ln=phat(1);sig2_ln=phat(2);
ln_pdf=1./(sqrt(2*pi*sig2_ln)*x).*exp(-(1/2)*(log(x)-mu_ln).^2/sig2_ln);
ln_pdf(isnan(ln_pdf))=zeros(size(find(isnan(ln_pdf))));%log of zero probs.
ln_hist=ln_pdf*dx*N;


phat=gamfit(rad);%MLE Weibull distribution (Stats toolbox)
phat=fminsearch('chi2_gam',phat);
al_g=phat(1);be_g=phat(2);
ga_pdf=gampdf(x,al_g,be_g);
ga_hist=ga_pdf*dx*N;

[xn,pn]=chi2fit(hgram,nor_hist,8,2);
[xl,pl]=chi2fit(hgram,ln_hist,8,2);
[xg,pg]=chi2fit(hgram,ga_hist,8,2);
```

```
function normalize, roifile, rad, FMT=fmt
;this function is designed to normalize all data in a selected roi
a = roi_extract('cr09m113_bip.cub', roifile, 'cr09m113_bip.hdr')
rad=total(a,1)
t = transpose(rad)
n = a/rebin(t, 210, n_elements(t))
return,a
end
```

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Devore, J. L., *Probability and Statistics for Engineering and the Sciences: Fourth Edition,* Brooks/Cole, 1995.

Goetz, A. F. H., "Imaging Spectrometry for Remote Sensing: Vision to Reality in 15 Years," *SPIE Proceedings*, vol. 2480, pp.2-13, 1995.

Goetz, A. F. H., Kindel, B., and Boardman, J. W., "Subpixel Target Detection in HYDICE Data from Cuprite, Nevada," *SPIE Proceedings*, vol. 2819, pp.7-14, 1996.

Healey, G. and Slater, D., "Models and Methods for Automated Material Identification in Hyperspectral Imagery Acquired Under Unknown Illumination and Atmospheric Conditions," *IEEE Transactions on Geoscience Remote Sensing*, vol. 37, pp. 2706-2717, 1999.

Hernandez-Baquero, Erich, "Survey of Advanced Technologies in Imaging Science for Remote Sensing." [http://www.cis.rit.edu/class/simg707/Web_Pages/Survey_report.htm]. 1997.

Kailey, W. F. and Illing, L., "Use of Hyperspectral Imagery for Broad Area Detection of Small Targets," *SPIE Proceedings*, vol. 2819, pp. 15-23, 1996.

Kruse, F. A., "Use of Airborne Imaging Spectrometer Data to Map Minerals Associated with Hydrothermally Altered Rocks in the Northern Grapevine Mountains, Nevada, and California," *Remote Sensing of the Environment*, vol. 24, pp. 31-51, 1988.

Math Works, Inc., *MATLAB Statistics Toolbox User's Guide*, Math Works, 1999.

Rickard, L. J., Basedow, R., Zalewske, E., Silverglate, P. R., Landers, M., "HYDICE: An airborne system for hyperspectral imaging," *SPIE Proceedings*, vol. 1937, pp. 173-179, 1993.

Stefanou, M. S., *A Signal Processing Perspective of Hyperspectral Imagery Analysis Techniques,* Master's Thesis, Naval Postgraduate School, Monterey, California, June 1997.

Stein, D. W. J., "A Robust Exponential Mixture Detector Applied to Radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 35, no.2, 1999.

Vane, G. and Goetz, A. F. H., "Terrestrial Imaging Spectroscopy," *Remote Sensing of the Environment*, vol. 24, pp. 1-29, 1988.

Vane, G., Green, R. O., Chrien, T. G., Enmark, H. T., Hansen, E.G., and Porter, W.M., "The airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sensing of the Environment*, vol.44, pp. 127-143, 1993.

BIBLIOGRAPHY

Boardman, J. W., "Analysis, Understanding, and Visualization of Hyperspectral Data as Convex Sets in *N*-Space," *SPIE Proceedings*, vol. 2480, pp.14-22, 1995.

Breckinridge, J. B., "Evolution of Imaging Spectrometry: Past, Present and Future," *SPIE Proceedings*, vol. 2819, pp.2-6, 1996.

Fukunaga, K., *Introduction to Statistical Pattern Recognition*, Academic Press, 1972.

Goetz, A. F. H., Vane, G., Solomon, J. E., and Rock, B. N., "Imaging Spectrometry for Earth Remote Sensing," *Science*, vol. 228, no.4704, pp. 1147-1153, 1985.

Landgrebe, D., "On the Analysis of Multispectral and Hyperspectral Data," *SPIE Proceedings*, vol. 2480, pp. 37-46, 1995.

Mazer, A. S., Martin, M., Lee, M, and Solomon, J. E., "Image Processing Software for Imaging Spectrometry Data Analysis," *Remote Sensing of the Environment*, vol. 24, pp. 201-210, 1988.

Miller, J. W. V., Farison, J. B., and Shin, Y., "Spatially Invariant Image Sequences," *IEEE Transactions on Image Processing*, vol. 1, no.2, pp. 148-161, 1992.

Mooney, J. M., "Angularly Multiplexed Spectral Imager," *SPIE Proceedings*, vol. 2480, pp. 65-77, 1995.

Price, J. C., "How Unique are Spectral Signatures?" *Remote Sensing of the Environment*, vol. 49, pp. 181-186, 1994.

Scharf, L. L., *Statistical Signal Processing: Detection, Estimation, and Time-Series Analysis*, Addison-Wesley, 1991.

Stein, D. W. J., "Detection of Small Targets in Clutter Modeled as a Gaussian Mixture or Hidden Markov Process," *SPIE Proceedings*, vol. 3163, pp.33-44, 1997.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST